

Hybrid Variational/Ensemble Data Assimilation

JCSDA Seminar,

World Weather Building, Washington DC, January 31st 2007

Dale Barker@, X. Wang\$, X. Y. Huang@, K. W. Chung#, Y.
Chen@, C. Snyder@, and A. C. Lorenc%

@ NCAR, # KMA, \$ NOAA, % Met. Office

Acknowledgements: AFWA, NSF, and KMA

Outline of Talk

- 1) Introduction
- 2) Examples of Flow-Dependent Errors in Variational DA.
- 3) 4D-Var for WRF: Status Report
- 4) Cycling WRF/ETKF Ensemble Prediction System.
- 5) Hybrid Variational/Ensemble Data Assimilation

1. Introduction



NCAR

Mesoscale & Microscale Meteorological Division

Motivation

- Intuitively expect forecast errors to be flow-dependent.
- 4D-Var permits flow-dependence, but background errors typically static.
- Ensemble DA implicitly flow-dependent, but issues (e.g. sampling error).
- Studies (e.g. Hamill and Snyder 2000, Etherton and Bishop 2004) show promising results of a hybrid variational/ensemble approach.
- Unification of variational/ensemble DA resources in a hybrid system?

Flow-Dependent EnKF Forecast Error Covariances

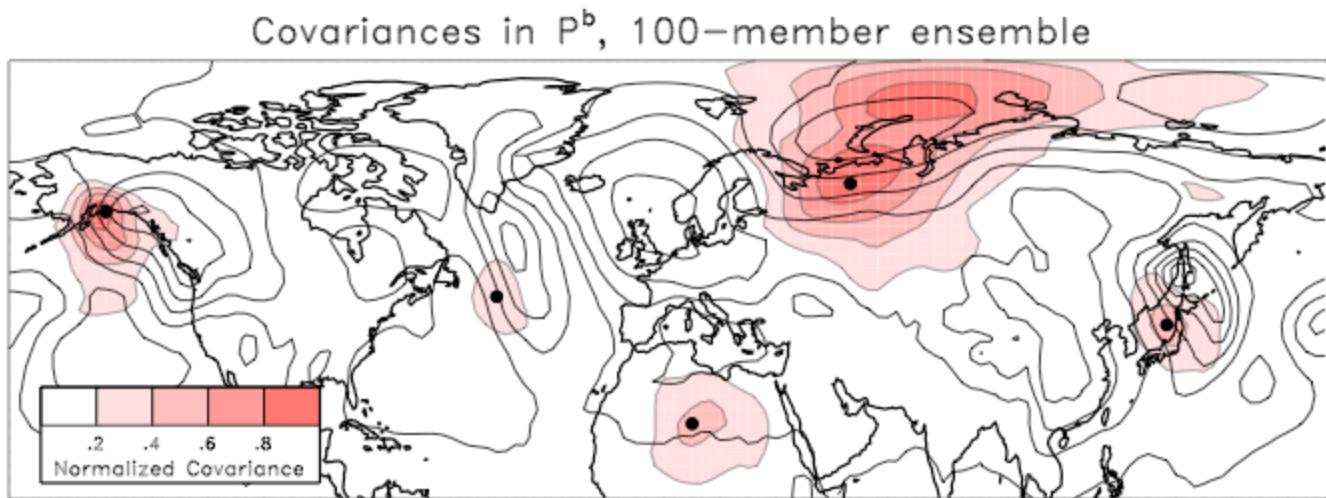


Figure 2. Background-error covariances (colors) of sea-level pressure in the vicinity of five selected observation locations, denoted by dots. Covariance magnitudes are normalized by the largest covariance magnitude on the plot. Solid lines denote ensemble mean background sea-level pressure contoured every 8 hPa.

From Hamill (2006)

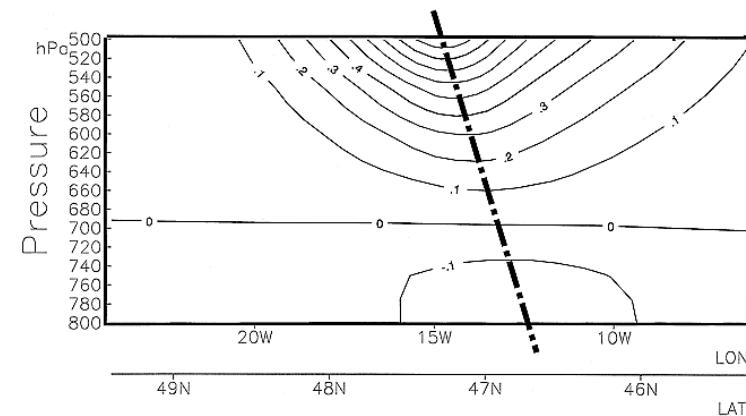
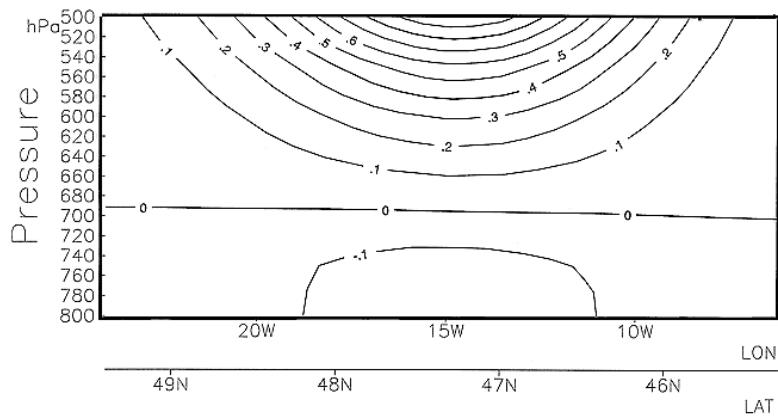
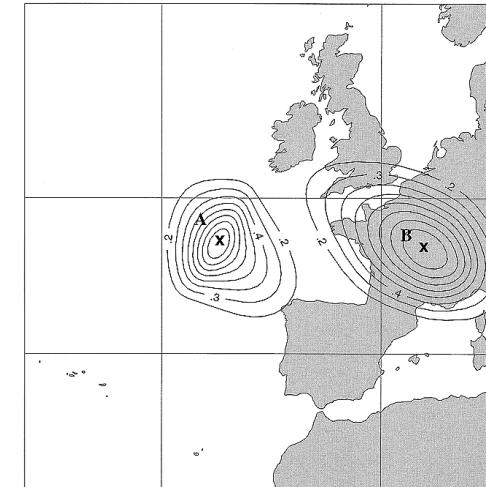
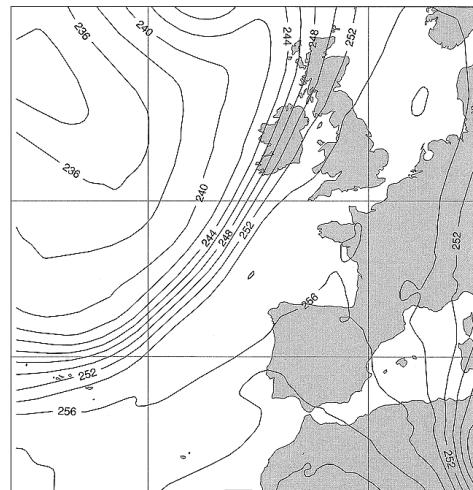
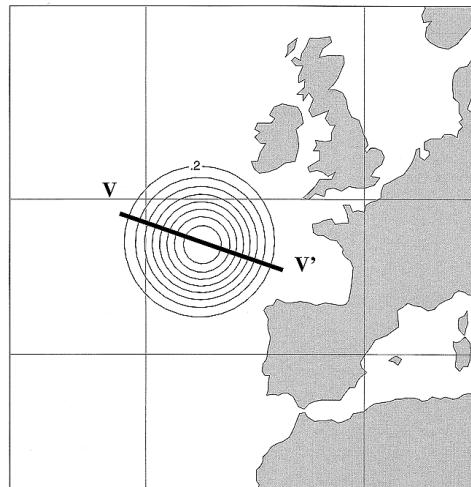


NCAR

Mesoscale & Microscale Meteorological Division

2. Examples Of Flow-Dependent Errors in Variational DA

Semi-Geostrophic Transformation



From Desroziers (1997)



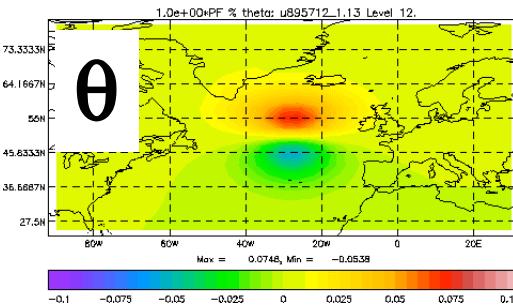
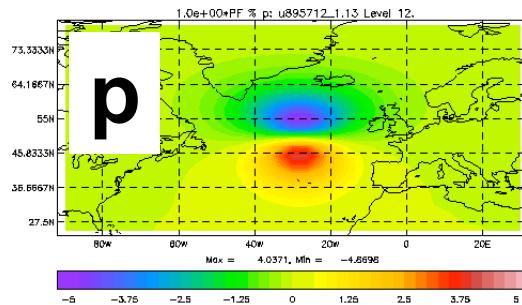
NCAR

Mesoscale & Microscale Meteorological Division

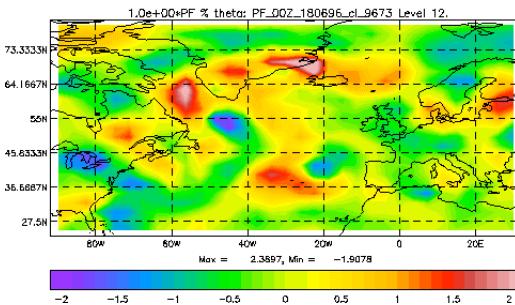
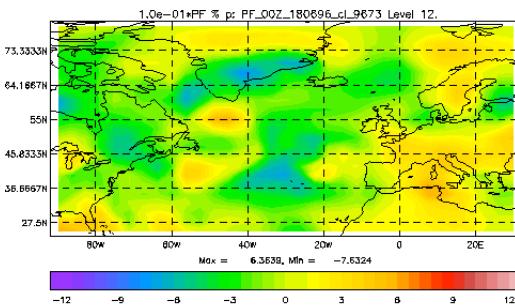
Flow-Dependence Via Extra Control Variables 1

- Increments due to a u-wind observation at 50N, 30W (O-B=1m/s)

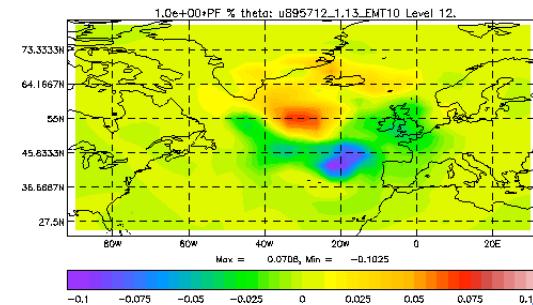
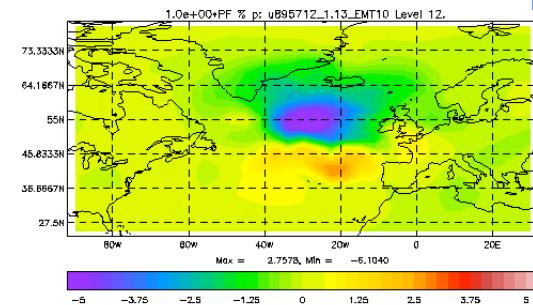
Static



Bred-Mode



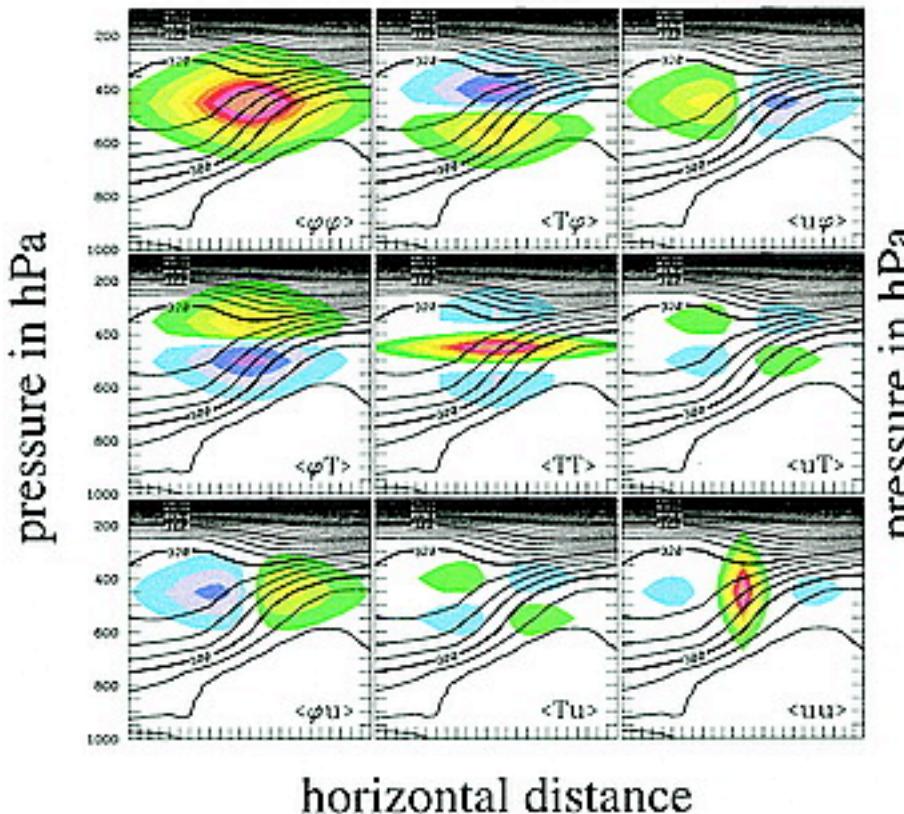
Flow-Dependent



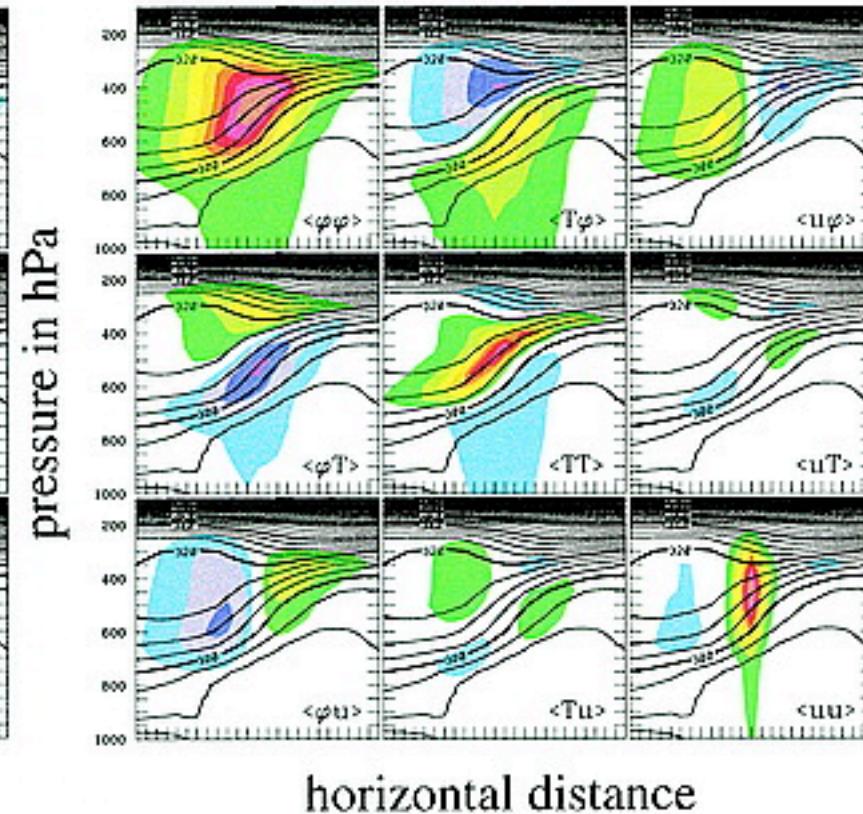
Barker (1999)

Observation-space 3D-Var (e.g. PSAS, NAVDAS)

(a) Pressure formulism



(b) Isentropic formulism



(from Daley and Barker 2000)



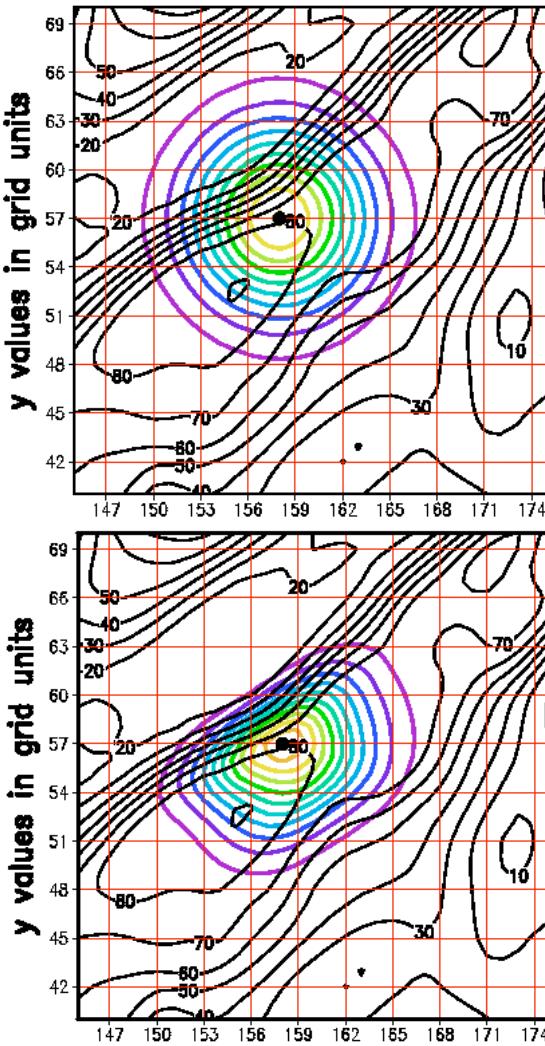
NCAR

Mesoscale & Microscale Meteorological Division

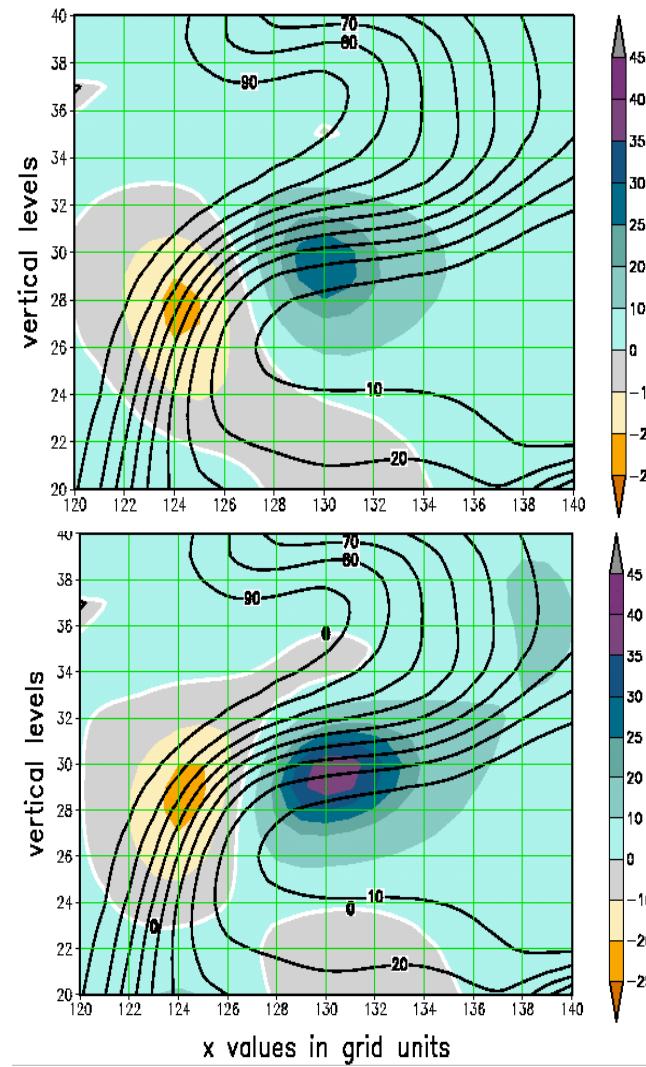
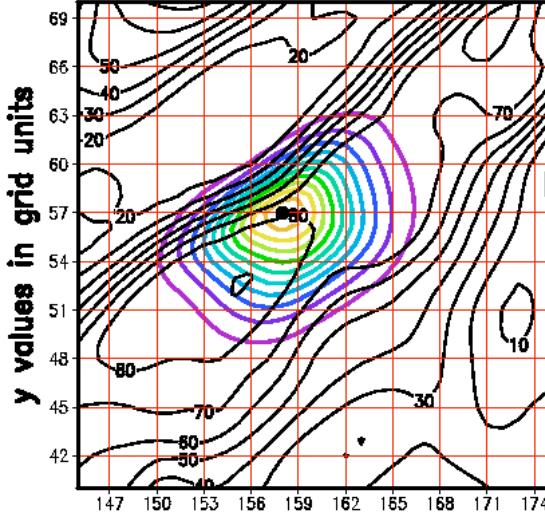
Flow-Dependence Via Anisotropic Recursive Filters

(courtesy NCEP/EMC)

Isotropic:



Anisotropic:

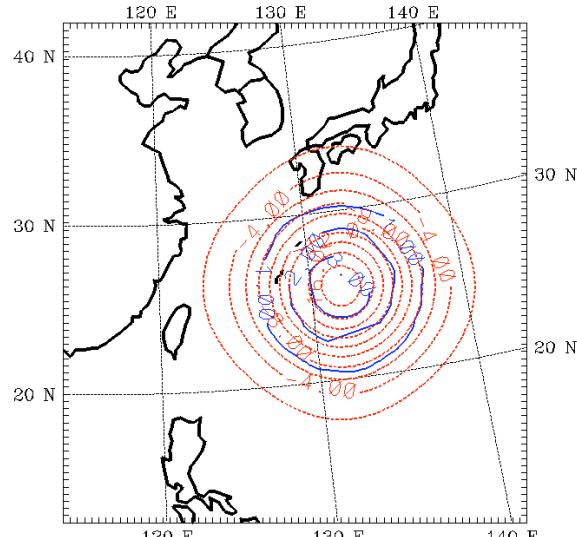


NCAR

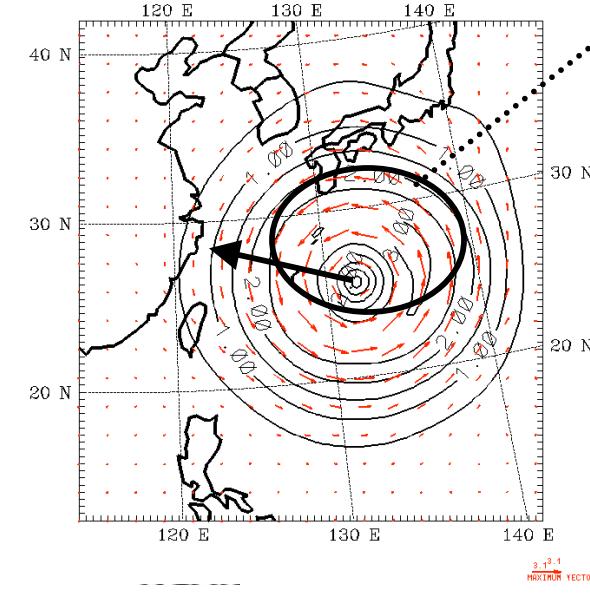
Mesoscale & Microscale Meteorological Division

Flow-Dependence Via Nonlinear (Dynamical) Balance Equation

$$\nabla^2 p'_b = -\nabla \cdot \bar{\rho} [\bar{v} \cdot \nabla v' + v' \cdot \nabla \bar{v} + f \mathbf{k} \times v']$$



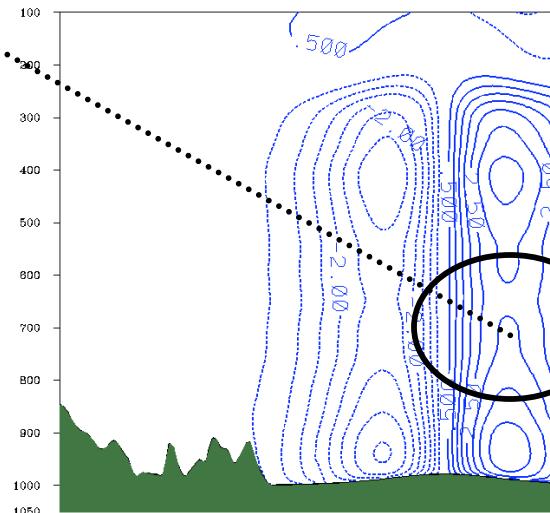
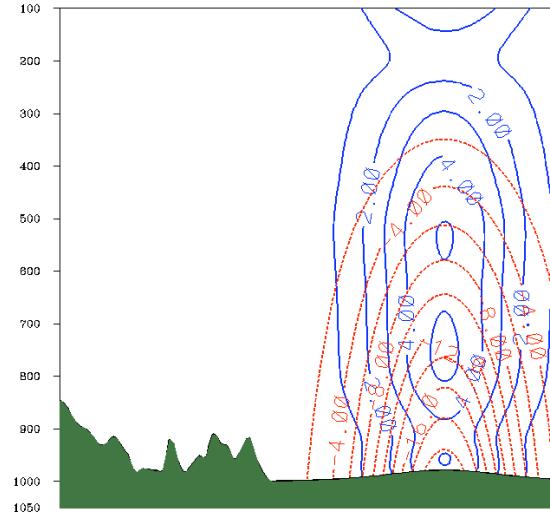
Pressure,
Temperature



Impact of
cyclostrophic balance
term

Wind Speed,
Vector,
v-wind component.

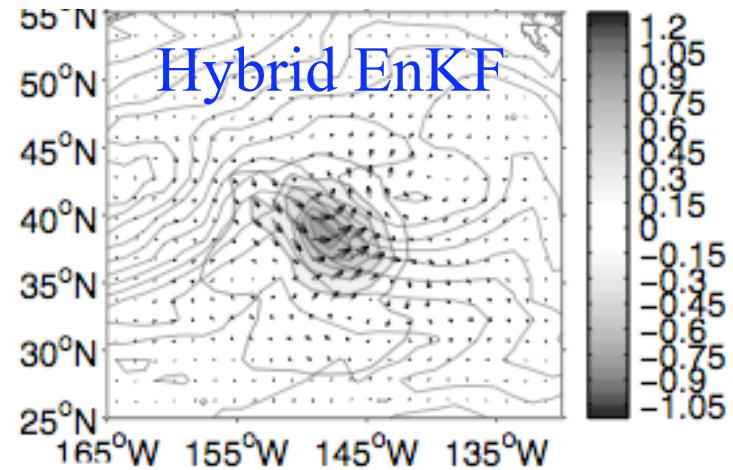
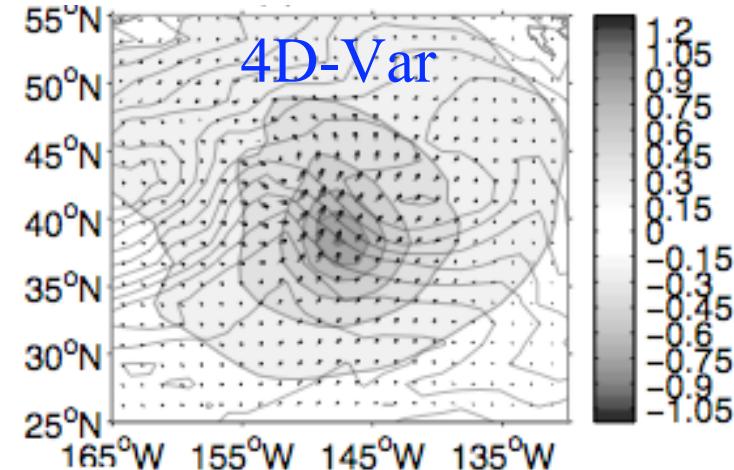
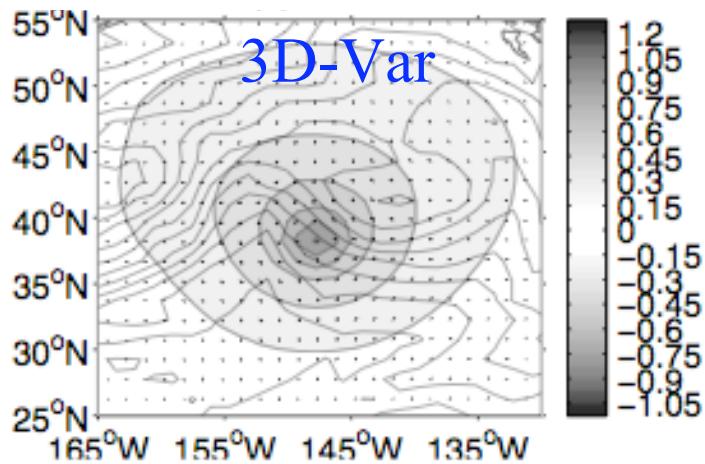
Barker et al (2004)



Mesoscale

Flow-Dependence Via Extra Control Variables 2

- Temperature increments of single temperature observation



Buehner (2005)



NCAR

Mesoscale & Microscale Meteorological Division

3. 4D-Var for WRF: Status Update



NCAR

Mesoscale & Microscale Meteorological Division

4D-Var Capability Of WRF-Var

Black – Pre-existing [$\mathbf{U} = \mathbf{B}^{1/2}$, $\mathbf{v}^n = \mathbf{U}^{-1} (\mathbf{x}^n - \mathbf{x}^{n-1})$, \mathbf{R} , H]

Green – Pre-existing, but modification required (\mathbf{H} , \mathbf{H}^T)

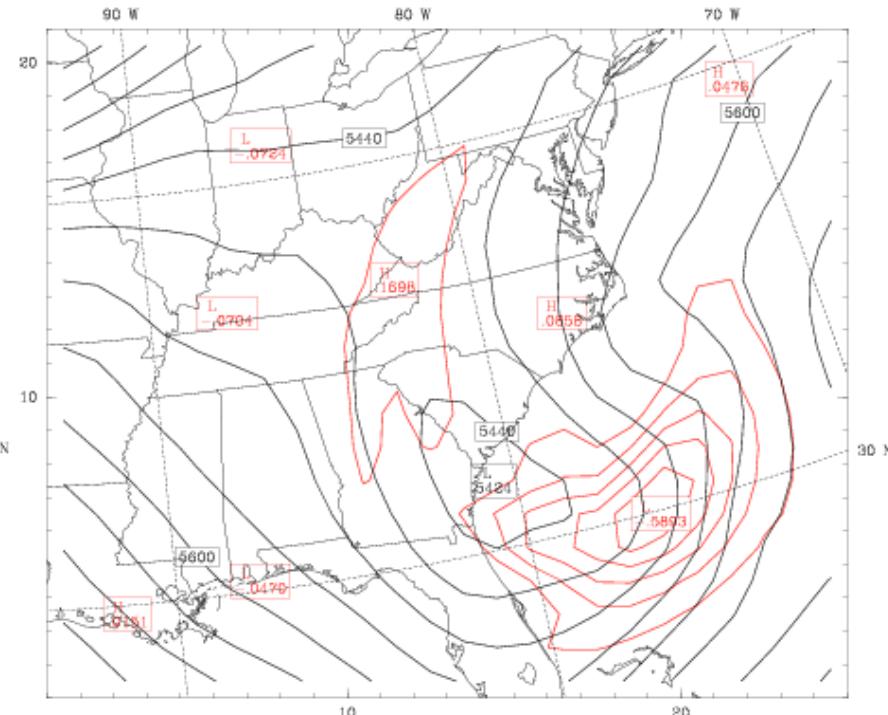
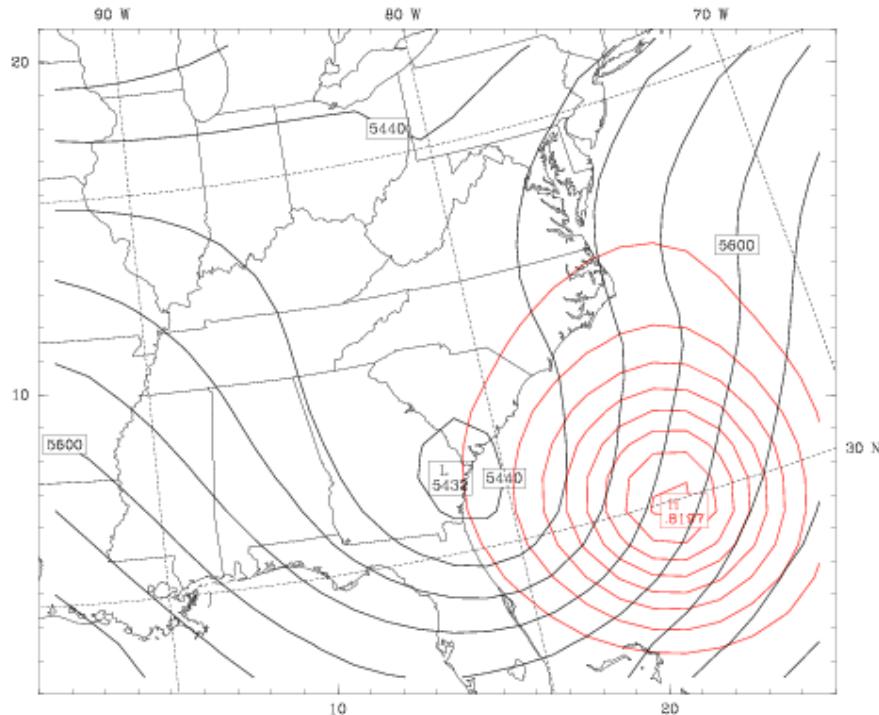
Blue – Pre-existing (M : WRF-ARW)

Red – new development (\mathbf{M} , \mathbf{M}^T , \mathbf{S}_{v-w} , \mathbf{S}_{w-v} , \mathbf{S}_{w-v}^T , \mathbf{S}_{v-w}^T)

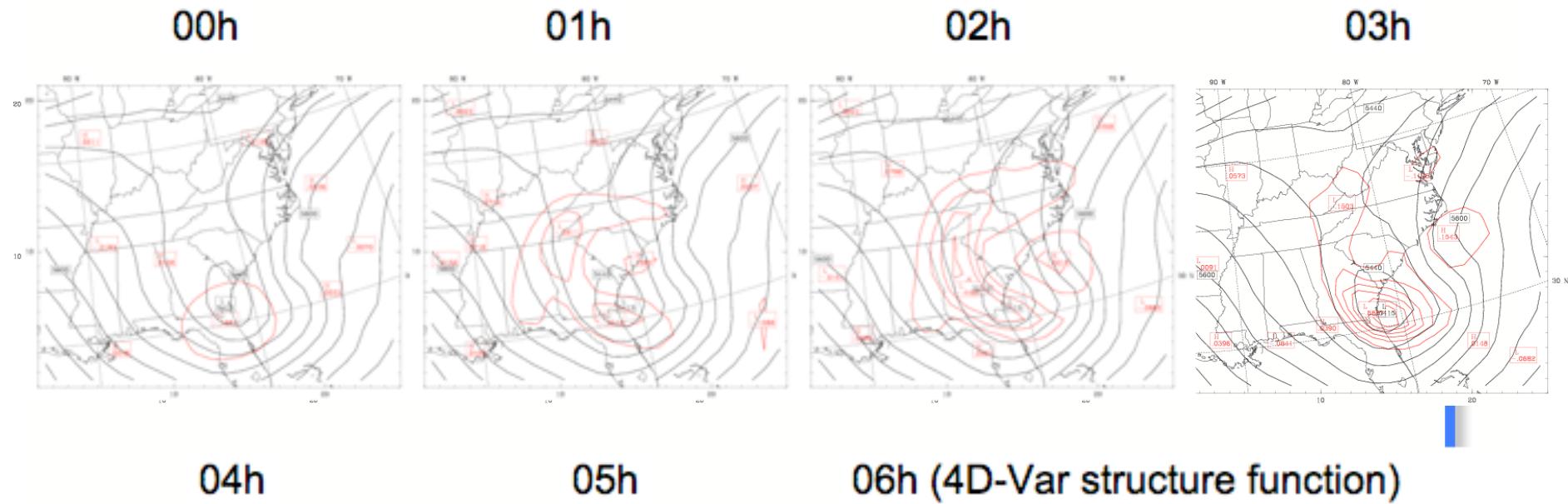
$$J'_{vn} = \mathbf{v}^n + \sum_{i=1}^{n-1} \mathbf{v}^i + \mathbf{U}^T \mathbf{S}_{v-w}^T \sum_{k=1}^K \mathbf{M}_k^T \mathbf{S}_{w-v}^T \mathbf{H}_k^T \mathbf{R}^{-1} [\mathbf{H}_k \mathbf{S}_{w-v} \mathbf{M}_k \mathbf{S}_{v-w} \mathbf{U}^{-1} \mathbf{v}^n + H_k(M_k(\mathbf{x}^{n-1})) - \mathbf{y}_k]$$

(Huang, et.al. 2006: Preliminary results of WRF 4D-Var. WRF users' workshop, Boulder, Colorado.
http://www.mmm.ucar.edu/wrf/users/workshops/WS2006/abstracts/Session04/4_5_Huang.pdf)

500mb θ increments from 3D-Var at 00h and from 4D-Var at 06h due to a 500mb T observation at 06h



500mb θ increments at 00,01,02,03,04,05,06h to a 500mb T ob at 06h

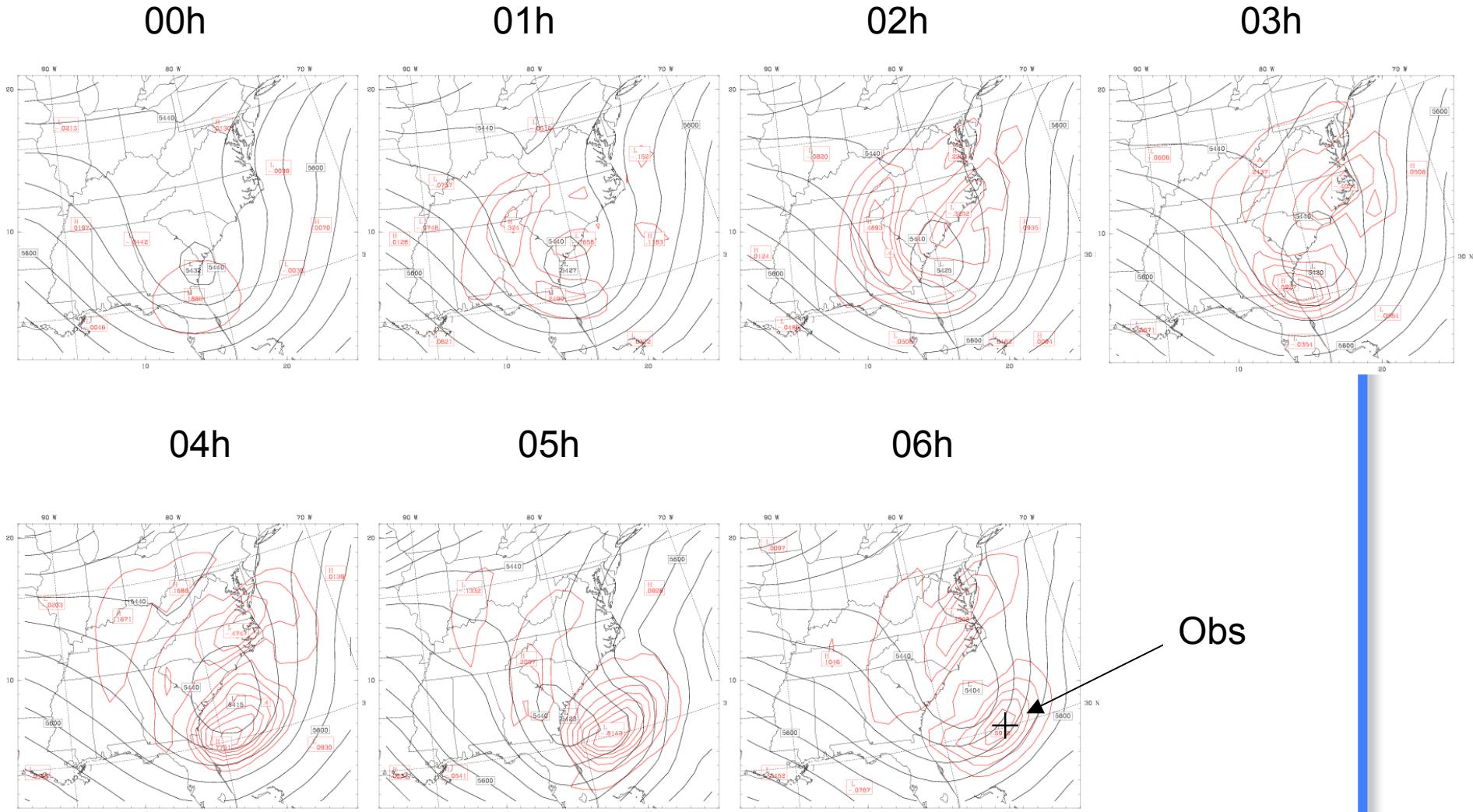


NCAR

Mesoscale & Microscale Meteorological Division

Obs

500mb θ difference at 00,01,02,03,04,05,06h between two nonlinear runs (one from background; one from 4D-Var)



NCAR

Mesoscale & Microscale Meteorological Division

Real Case: Typhoon Haitang Experimental Design

- Domain configuration: 91x73x17, 45km
- Observations from Taiwan CWB operational database.
- Experiments are conducted before Haitang's landfall at 0000 UTC 18 July 2005.
 - **FGS** – forecast from the background [The background fields are 6-h WRF forecasts from National Center for Environment Prediction (NCEP) GFS analysis.]
 - **AVN** – forecast from the NCEP AVN analysis
 - **3DVAR** – forecast from 3D-Var (NoFGAT)
 - **4DVAR** – forecast from 4D-Var

Observations used in 4DVAR at 0000UTC 16 July 2005

	u	v	T	p	q	dZ
TEMP	727	724	869		697	
TEMPSurf	6	8	8	8	8	
SYNOP	199	218	237	226	236	
SATOB	3187	3182				
AIREP	923	930	939			
PILOT	159	160				
METAR	167	191	216	0	200	
SHIP	69	70	77	79	73	
SATEM						511
BUOY	67	67	0	64	0	
BOGUS	1200	1200	788	788	80	

(At 0600UTC 16 July: GPS refractivity 2594, QuikScat u 2594, v 2605)



NCAR

Mesoscale & Microscale Meteorological Division

The track error in km averaged over 48 h

Time	FGS	AVN	3DREF	4DREF
1512	84	82	73	66
1518	82	130	71	85
1600	138	83	92	68
1606	92	83	77	78
1612	96	90	74	61
1618	95	67	101	96
1700	100	86	88	84
1706	111	104	97	116
1712	126	134	131	133
1718	144	126	126	127
1800	150	159	169	156
Average	110.7	104.0	99.9	97.3



NCAR

Current status of 4D-Var Development

- Modifications to 3D-Var aspects of WRF-Var complete.
- WRF-ARW TL and adjoint models developed (**TAF**).
- Prototype (parallel, but limited physics) undergoing tests.
- Major 2007 tasks: Optimization, more physics, testing.
- Implementation at AFWA planned for 2008.

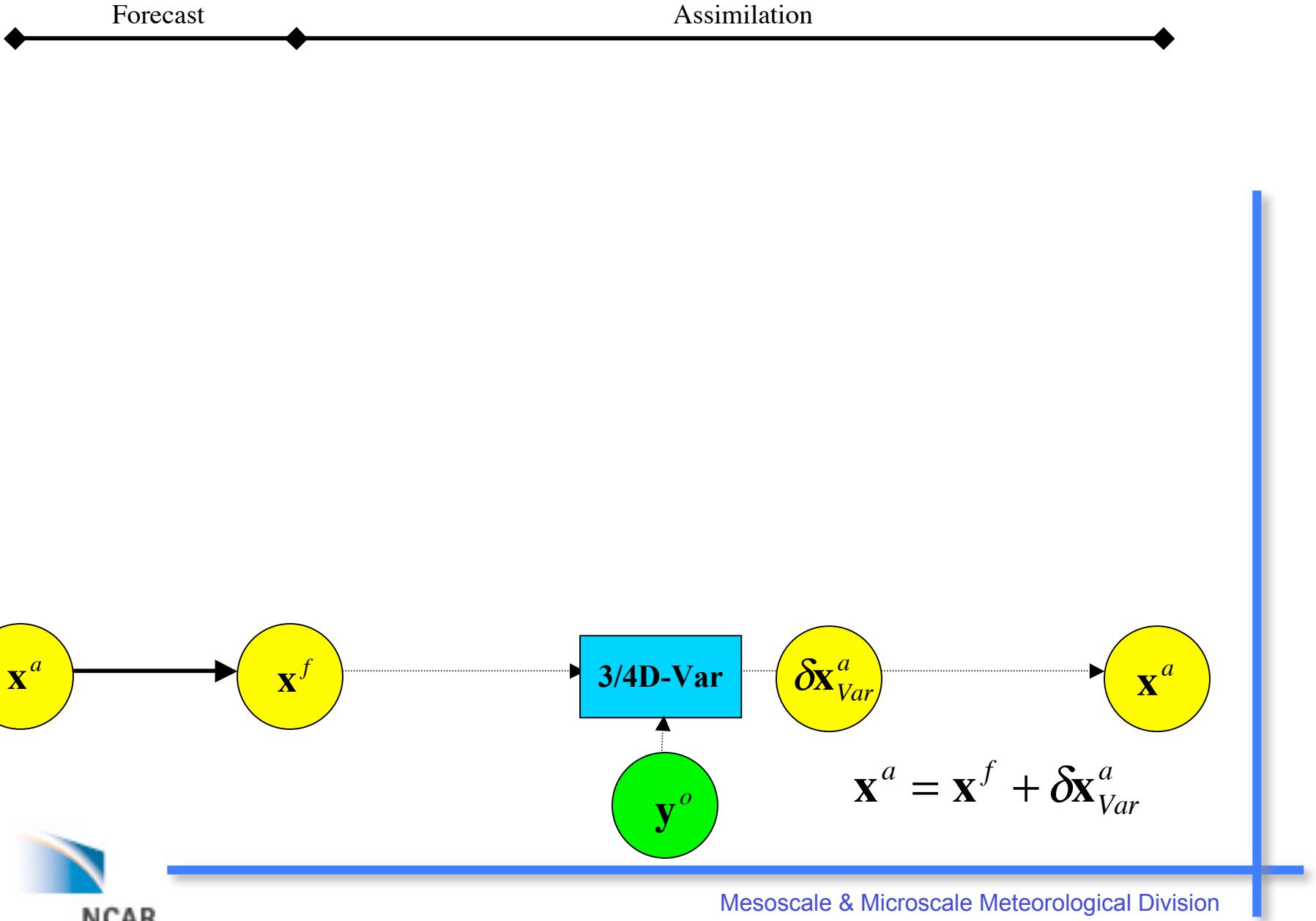
4. Cycling WRF/WRF-Var/ETKF



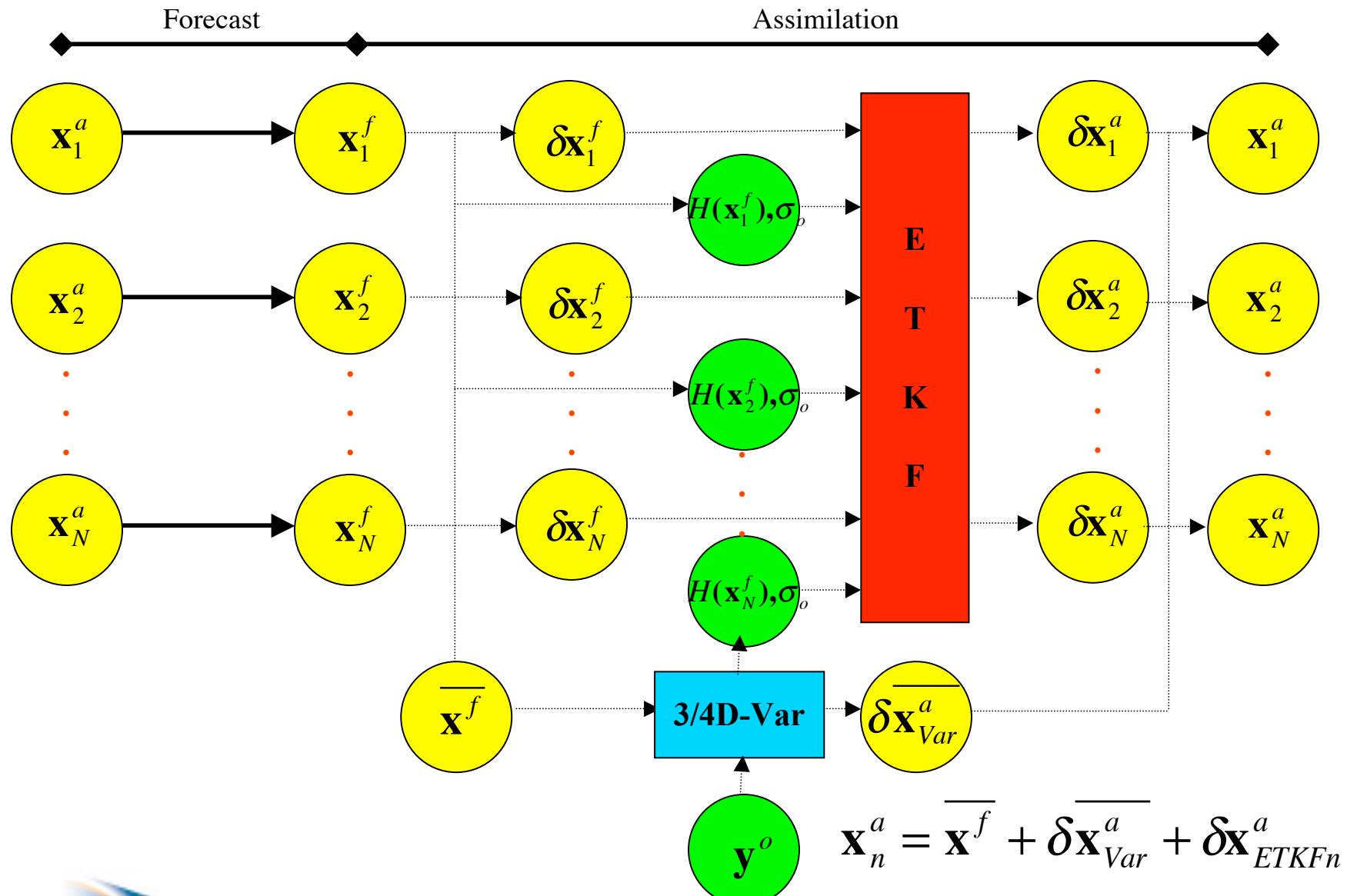
NCAR

Mesoscale & Microscale Meteorological Division

1. Deterministic Cycling NWP System



Cycling WRF/WRF-Var/ETKF System



NCAR

Mesoscale & Microscale Meteorological Division

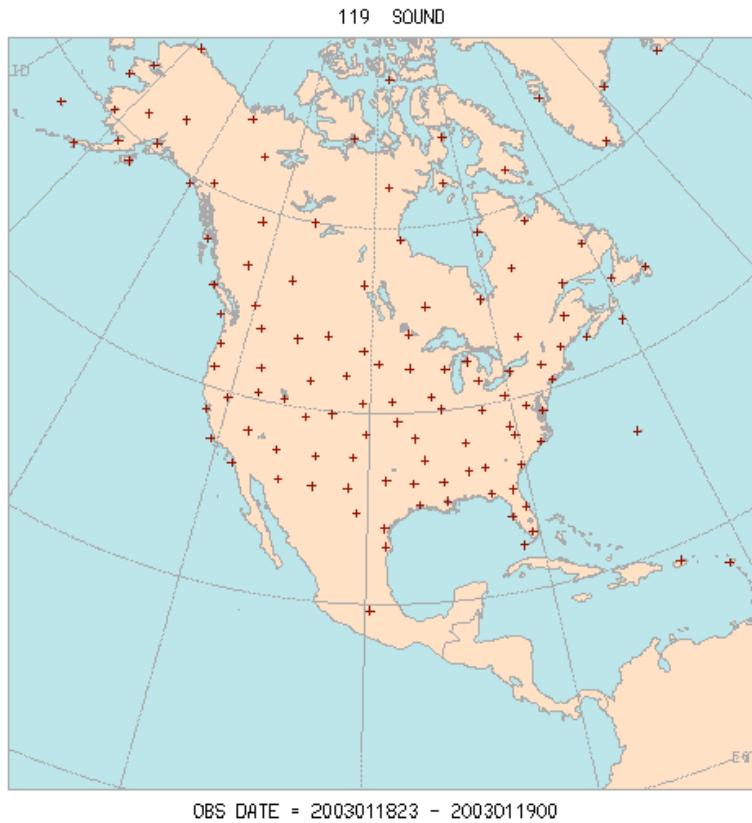
The Ensemble Transform Kalman Filter (ETKF)

- The ETKF (Bishop et al 2001) is an example of a deterministic filter.
- The ETKF finds the transformation matrix \mathbf{T} so that

$$\mathbf{P}_{ens}^f = \frac{1}{N_e - 1} (\mathbf{X}'^f \mathbf{T}) (\mathbf{X}'^f \mathbf{T})^T$$

- ETKF is fast (matrices are computed in ensemble space).
- The ETKF cannot localize so particularly prone to sampling error effects.
- Update perturbations $\mathbf{X}'^a = \mathbf{X}'^f \mathbf{T}$, but NOT ensemble mean.
- Use ETKF for EPS IC perts., and flow-dependent errors in hybrid DA.

Domain For Initial WRF/ETKF Tests

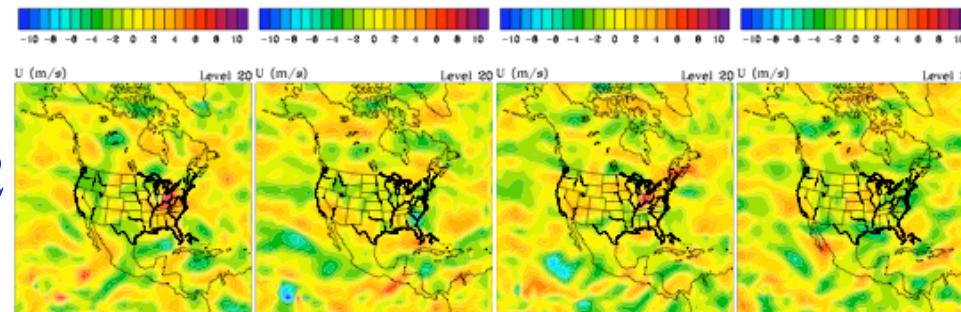
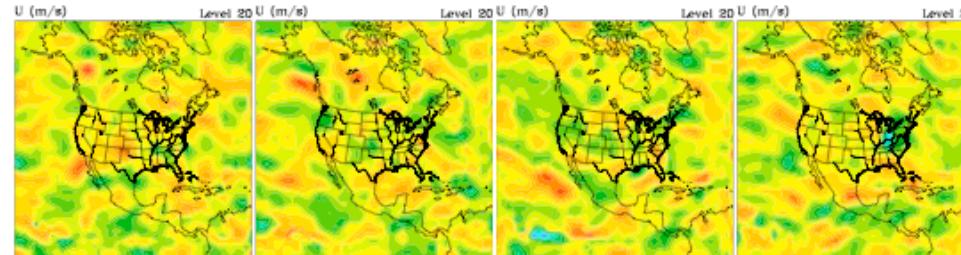


- Low-resolution (200km) CONUS domain.
- Same domain used for initial 3D-Var/EnKF comparison.
- January 2003 test period.
- Assimilate sondes at 12hr intervals.
- 30/50 member ETKF.
- Sampled 3D-Var covariances used to provide initial/lateral boundary perturbations.

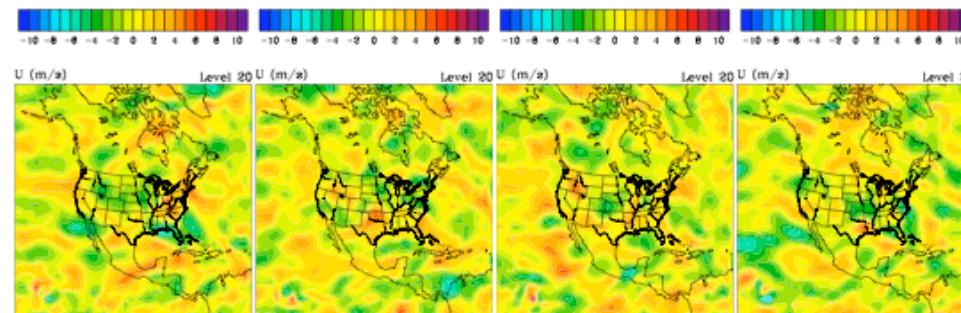
ETKF Ensemble Perturbations (Members 1-12)

2003010100 T+12

U Level 20



Valid:2003010112



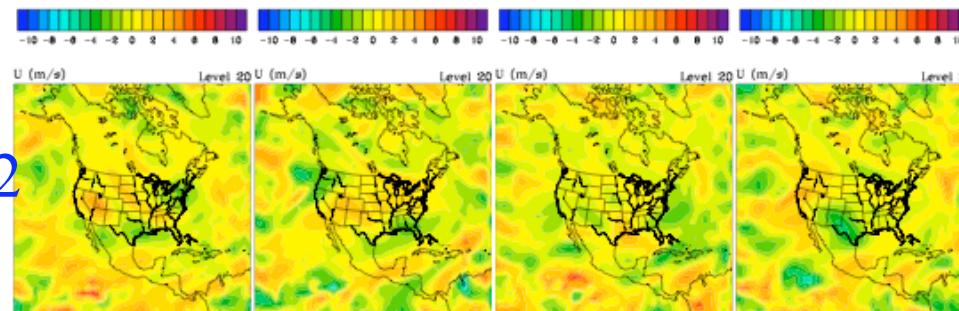
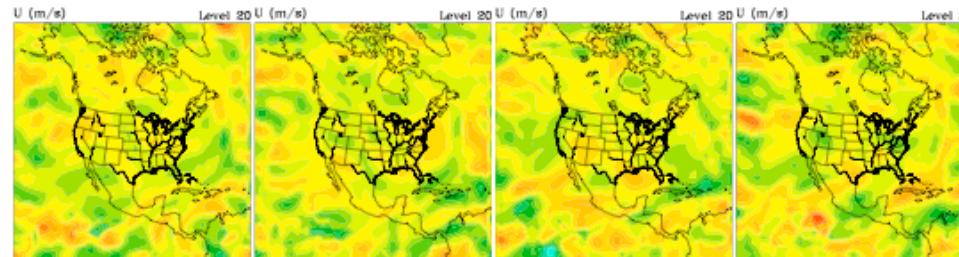
NCAR

Mesoscale & Microscale Meteorological Division

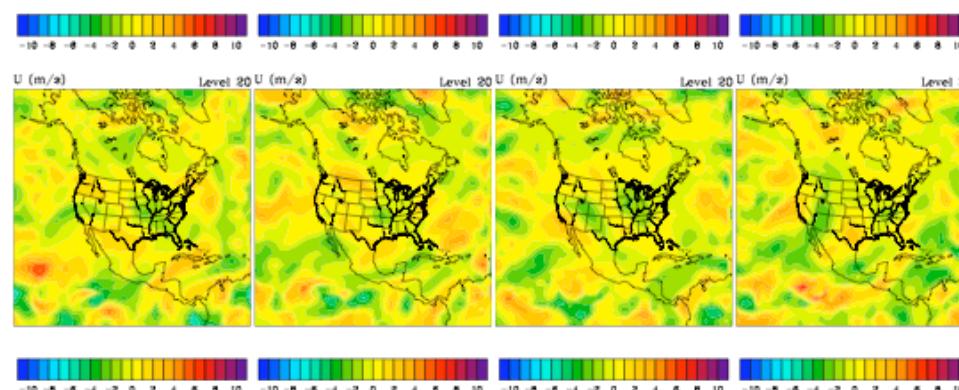
ETKF Ensemble Perturbations (Members 1-12)

2003011400 T+12

U Level 20



Valid:2003011412



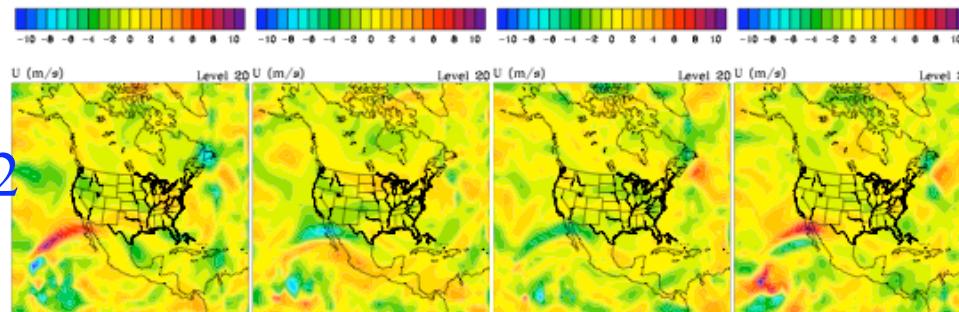
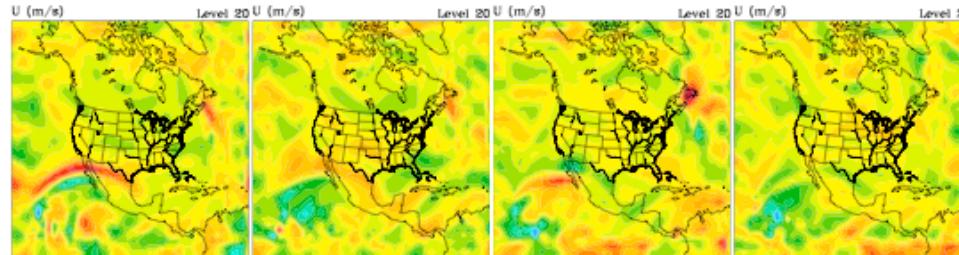
NCAR

Mesoscale & Microscale Meteorological Division

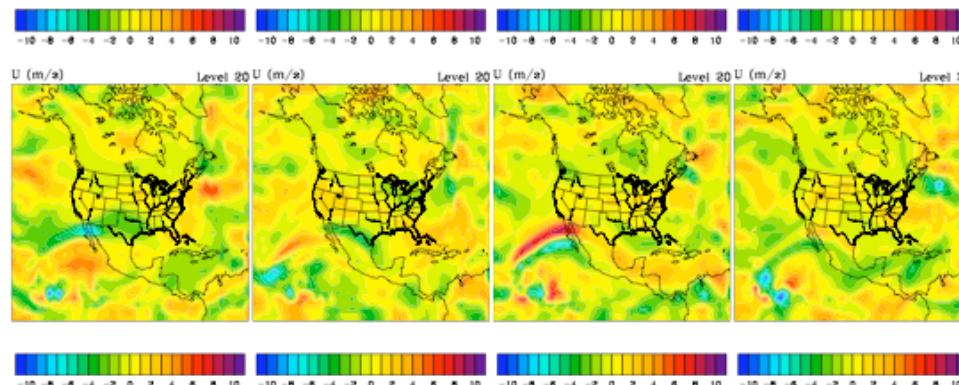
ETKF Ensemble Perturbations (Members 1-12)

2003012800 T+12

U Level 20



Valid:2003012812



NCAR

Mesoscale & Microscale Meteorological Division

ETKF Covariance Inflation

- Small ensemble size results in underestimate of analysis error covariances.
- Following Wang and Bishop (2003), boost ETKF ensemble spread via adaptive inflation factor $\tilde{\Pi}^{1/2}$

$$\mathbf{X}'^a = \mathbf{X}'^f \mathbf{T} \tilde{\Pi}^{1/2}$$

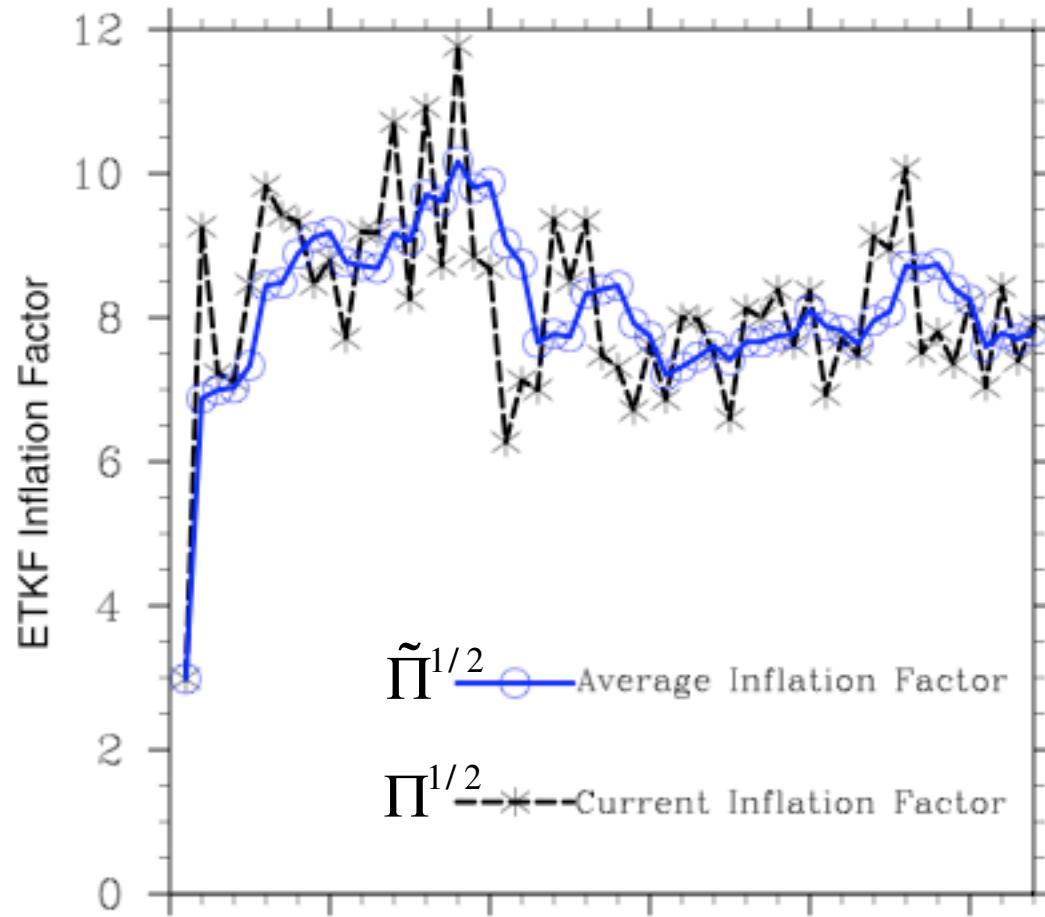
- Adaptively inflate at time i by matching spread to innovation vectors:

$$\Pi_i = (\tilde{\mathbf{d}}^T \cdot \tilde{\mathbf{d}} - N_o)_i / \sum_{n=1}^{N_e-1} \lambda_{ni} \quad \tilde{\Pi}_i = 1/I \sum_{j=i-I+1}^i \Pi_j$$

$$\tilde{\mathbf{H}} = \mathbf{H} / \sigma_o \quad \tilde{\mathbf{d}} = [\mathbf{y}^o - \overline{H(\mathbf{x}^f)}] / \sigma_o$$

- I is the averaging period, λ are the eigenvalues of $1/(N_e - 1) \tilde{\mathbf{H}} \mathbf{P}^f \tilde{\mathbf{H}}$

Convergence Of ETKF Covariance Inflation

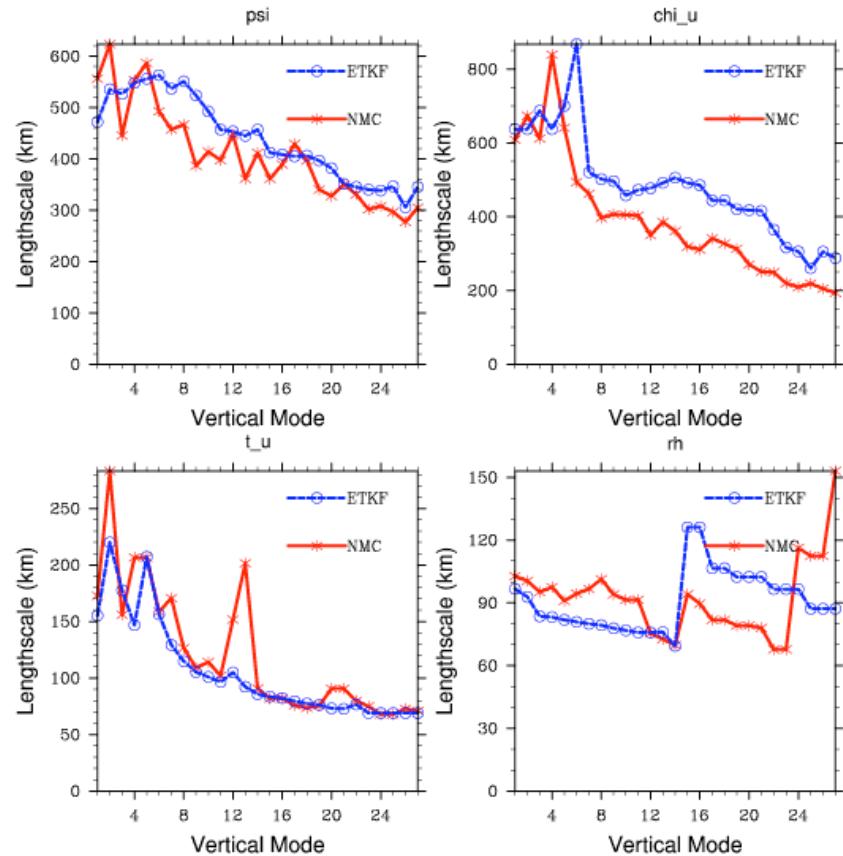


NCAR

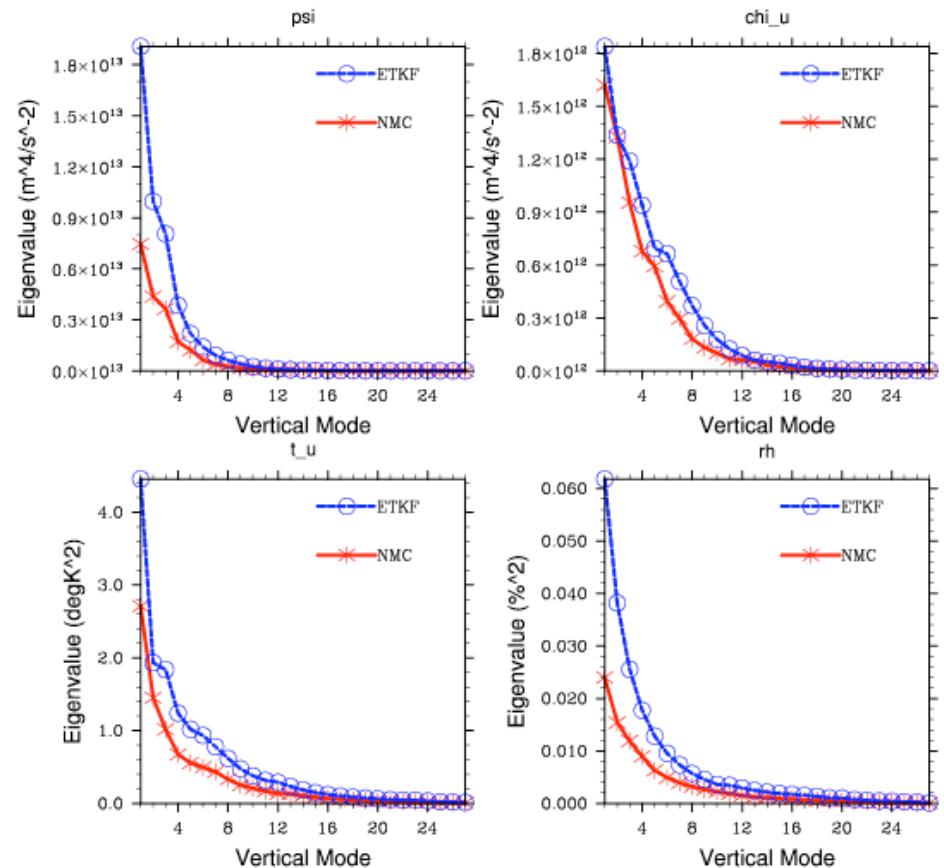
Mesoscale & Microscale Meteorological Division

ETKF vs NMC-method Climatological Covariances

Horizontal Correlation Scales



Eigenvalues (Variances)



ETKF variances implicitly tuned, unlike NMC-method



NCAR

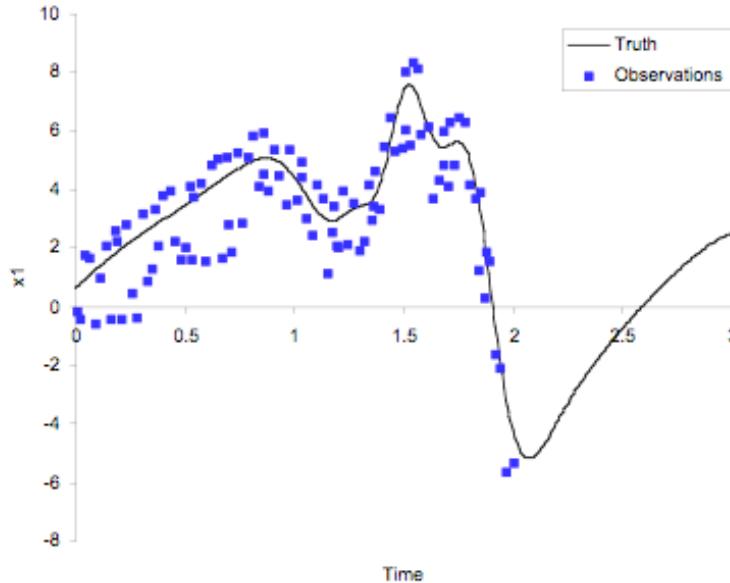
Mesoscale & Microscale Meteorological Division

5. Hybrid Variational/Ensemble Data Assimilation

Hybrid Testing With Lorenz 1996 model

(K. Y. Chung)

$$\frac{d\mathbf{x}_i}{dt} = (\mathbf{x}_{i+1} - \mathbf{x}_{i-2})\mathbf{x}_{i-1} - \mathbf{x}_i + \mathbf{F}$$



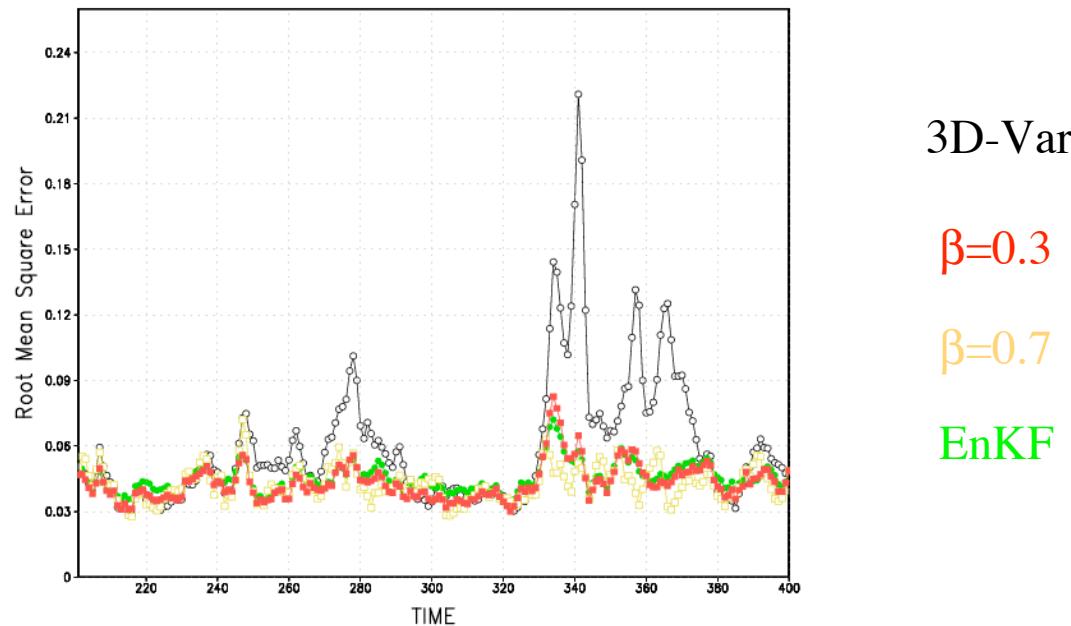
- 40-variable model ($i=1, 40$). $\mathbf{F}=8$.
- Periodic boundary conditions $\mathbf{x}_1 = \mathbf{x}_{40}$
- OSSE: $dt=6h$. Simulate obs every 12h. 400d run, verify last 200d.
- Hybrid 3D-Var/EnKF (Hamill and Snyder 2000): $\mathbf{P}^f = (1-\beta)\mathbf{P}_e^f + \beta\mathbf{B}_0$
- Flow-dependence \mathbf{P}_e^f via a) Lagged forecast diff., b) EnKF perturbations.

Lorenz Model OSSE Analysis Error, N=10

Root Mean Square Analysis Error (x100)

β	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
3D-Var						5.70			
Hy(FD)	7.06	5.23	4.88	4.78	4.27	4.24	4.40	4.45	4.44
Hy(EnKF)	4.21	4.33	4.19	4.46	4.32	4.37	4.86	5.00	5.20
EnKF					4.59				

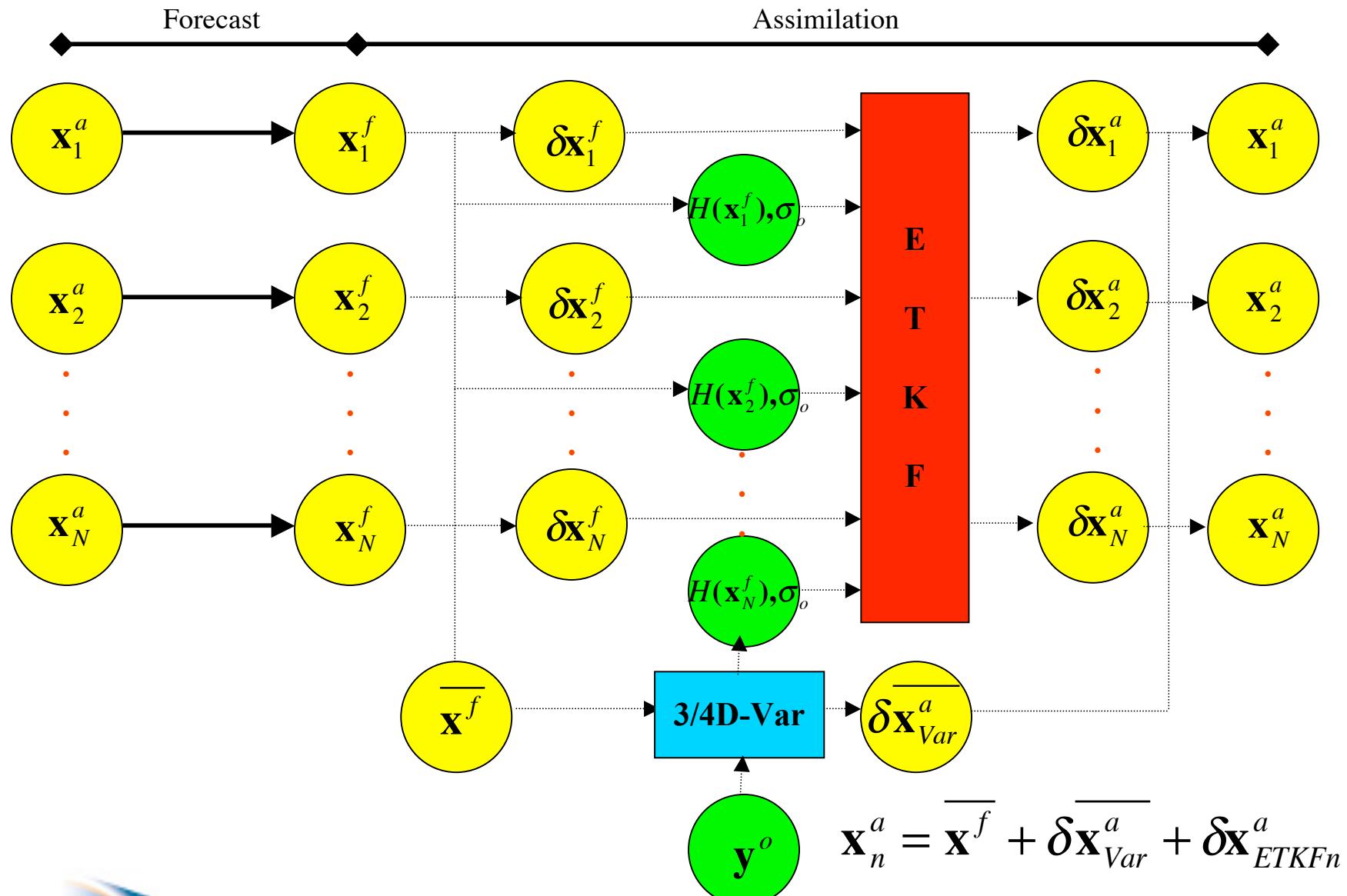
Optimal $\beta=0.3$
consistent with
Etherton and
Bishop (2004)



NCAR

Mesoscale & Microscale Meteorological Division

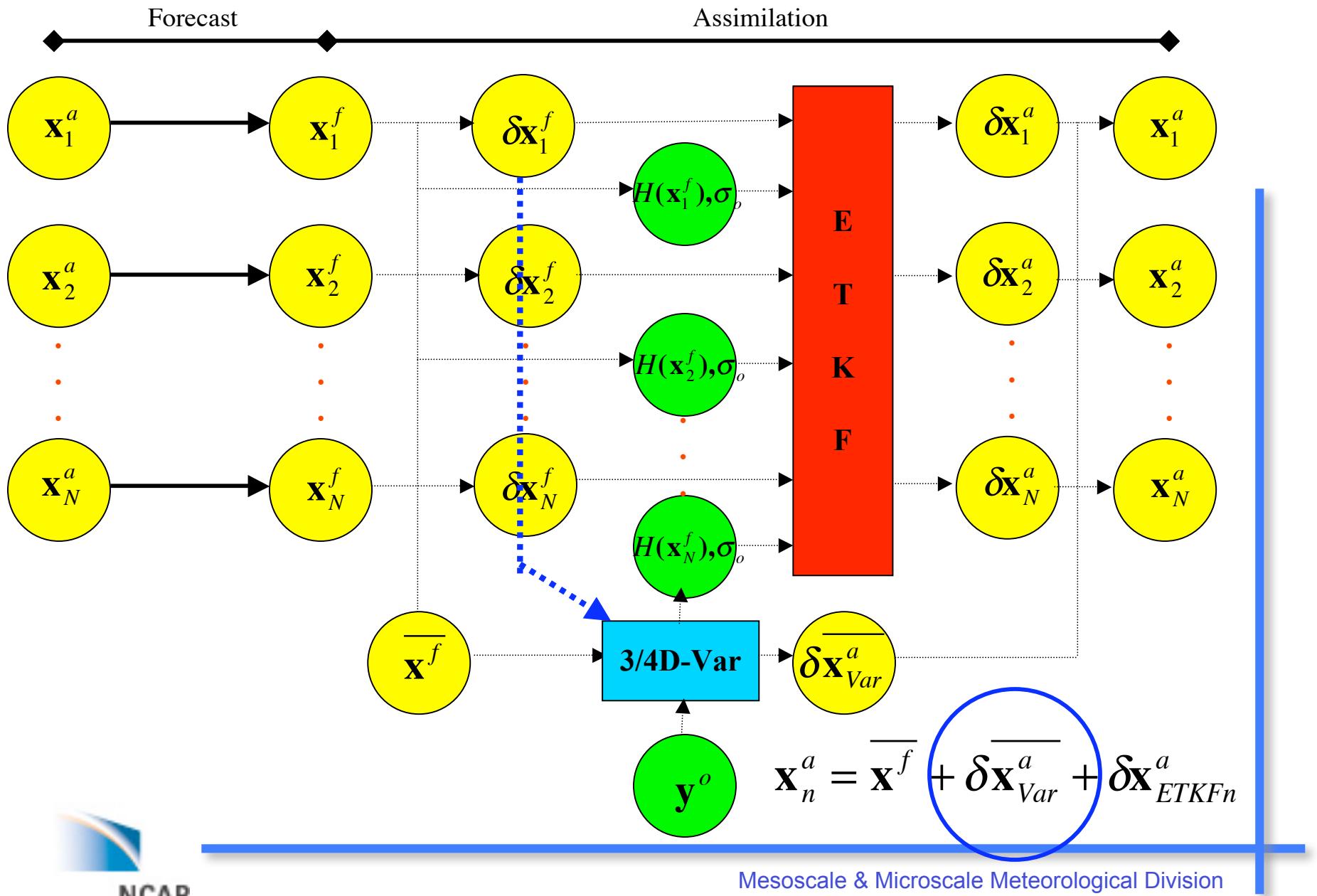
Cycling WRF/WRF-Var/ETKF System



NCAR

Mesoscale & Microscale Meteorological Division

Cycling WRF/WRF-Var/ETKF System (Hybrid DA)



NCAR

Mesoscale & Microscale Meteorological Division

Incremental J_b Preconditioning (Example: MetO, WRF-Var)

$$J_b[\delta\mathbf{x}(t_0)] = \frac{1}{2} \left\{ \delta\mathbf{x}(t_0) - [\mathbf{x}^b(t_0) - \mathbf{x}^g(t_0)] \right\}^T \mathbf{B}_o^{-1} \left\{ \delta\mathbf{x}(t_0) - [\mathbf{x}^b(t_0) - \mathbf{x}^g(t_0)] \right\}$$

- Define **preconditioned control variable** v space transform

$$\delta\mathbf{x}(t_0) = \mathbf{U}\mathbf{v}$$

where \mathbf{U} transform **CAREFULLY** chosen to satisfy $\mathbf{B}_o = \mathbf{U}\mathbf{U}^T$.

- Choose (at least assume) control variable components with uncorrelated errors:

$$J_b[\delta\mathbf{x}(t_0)] = \frac{1}{2} \sum_n v_n^2$$

- where $n \sim$ number pieces of independent information.

WRF-Var Background Error Modeling

cv_options		2 (original MM5)	3(GSI)	4 (Global)	5(regional)
Analysis increments	\mathbf{x}'	$u', v', T', q', p_s'(i, j, k)$			
Change of Variable	U_p	$\psi', \chi', p_u', q'(i, j, k)$	$\psi', \chi_u', T_u', \tilde{r}', p_{su}'(i, j, k)$		
Vertical Covariances	U_v	$\mathbf{B} = \mathbf{E} \Lambda \mathbf{E}^T$	RF	$\mathbf{B} = \mathbf{E} \Lambda \mathbf{E}^T$	
Horizontal Correlations	U_h	RF	Spectral	RF	
Control Variables	\mathbf{v}	$\mathbf{v}(i, j, m)$	$\mathbf{v}(l, n, m)$	$\mathbf{v}(i, j, m)$	

$$\delta\mathbf{x}(t_0) = \mathbf{U}\mathbf{v} = \mathbf{U}_p \mathbf{U}_v \mathbf{U}_h \mathbf{v}$$

Define control variables:

$$\psi'$$

$$r' = q'/q_s(T_b, q_b, p_b)$$

$$\chi' = \chi_u' + \chi_b'(\psi')$$

$$T' = T_u' + T_b'(\psi')$$

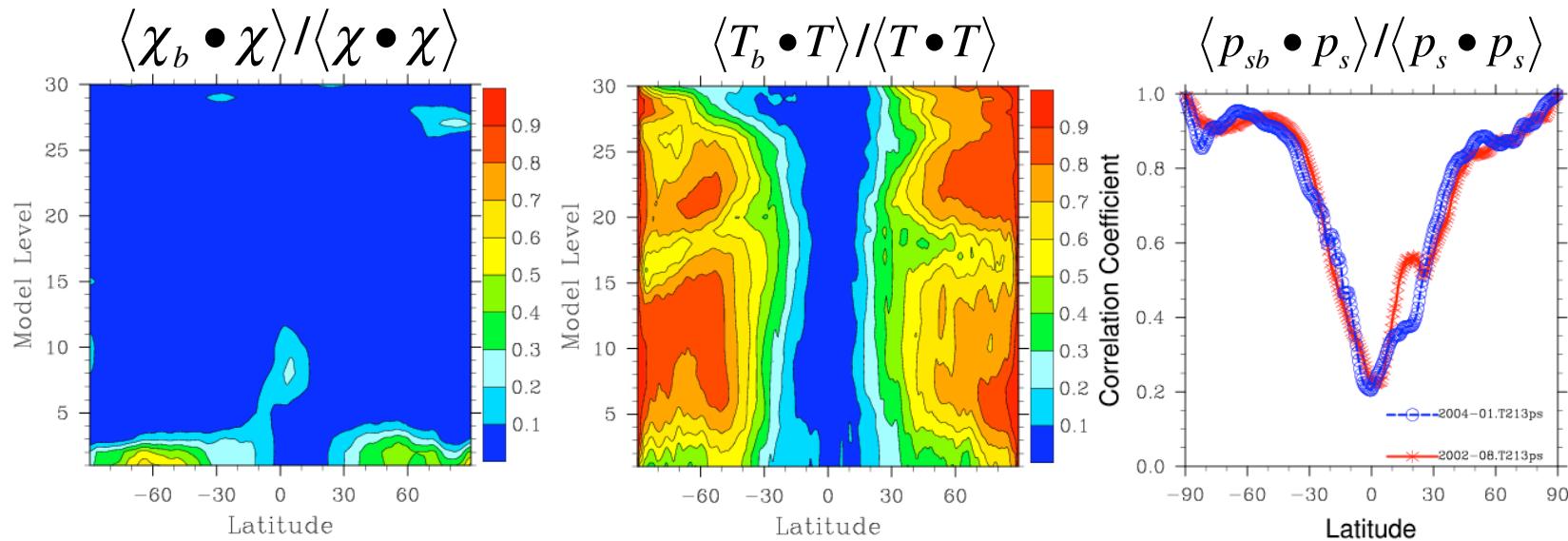
$$p_s' = p_{su}' + p_{sb}'(\psi')$$

WRF-Var Statistical Balance Constraints

- Define statistical balance after Wu et al (2002):

$$\chi'_b = c\psi' \quad T'_b(k) = \sum_{k1} G(k, k1)\psi'(k1) \quad p'_{sb} = \sum_k W(k)\psi'(k)$$

- How good are these balance constraints? Test on KMA global model data. Plot correlation between “Full” and balanced components of field:



NCAR

Mesoscale & Microscale Meteorological Division

Hybrid DA Via Additional Control Variables

- Define the matrix of ensemble perturbations as

$$\delta\mathbf{X}_f = (\delta\mathbf{x}_{f1}, \delta\mathbf{x}_{f2}, \dots, \delta\mathbf{x}_{fN}) \quad (1)$$

- Hybrid 3/4D-Var analysis increments give by

$$\delta\mathbf{x}_0 = \delta\mathbf{x}_{0d} + \delta\mathbf{X}_f \bullet \mathbf{a} \quad (2)$$

- Note flow-dependence $\delta\mathbf{X}_f$ constrained by a new set of control variables

$$\mathbf{a}^T = (\alpha_1, \alpha_2, \dots, \alpha_N) \quad (3)$$

- Could alternatively define the hybrid in control variable space, e.g.

$$\delta\psi_0 = \delta\psi_{0d} + \delta\psi_f \bullet \mathbf{a} \quad \delta\chi_{u0} = \delta\chi_{u0d} + \delta\chi_{uf} \bullet \mathbf{a} \quad (4)$$

- (4) better than (2) more when balance well known.

Cost Functions and Variance Conservation

- Flow-dependence is constrained by an additional cost-function J_α , i.e.

$$J = \frac{W_b}{2} \delta\mathbf{x}_0^T \mathbf{B}_o^{-1} \delta\mathbf{x}_0 + \boxed{\frac{W_\alpha}{2} \mathbf{a}^T \mathbf{A}^{-1} \mathbf{a} + \frac{1}{2} \sum_{i=0}^n \left[\mathbf{H}_i \delta\mathbf{x}(t_i) - \mathbf{d}_i \right]^T \mathbf{R}_i^{-1} \left[\mathbf{H}_i \delta\mathbf{x}(t_i) - \mathbf{d}_i \right]}$$

- Define empirical alpha covariance matrix $\mathbf{A} = \sigma_\alpha^2 \mathbf{A}_c$

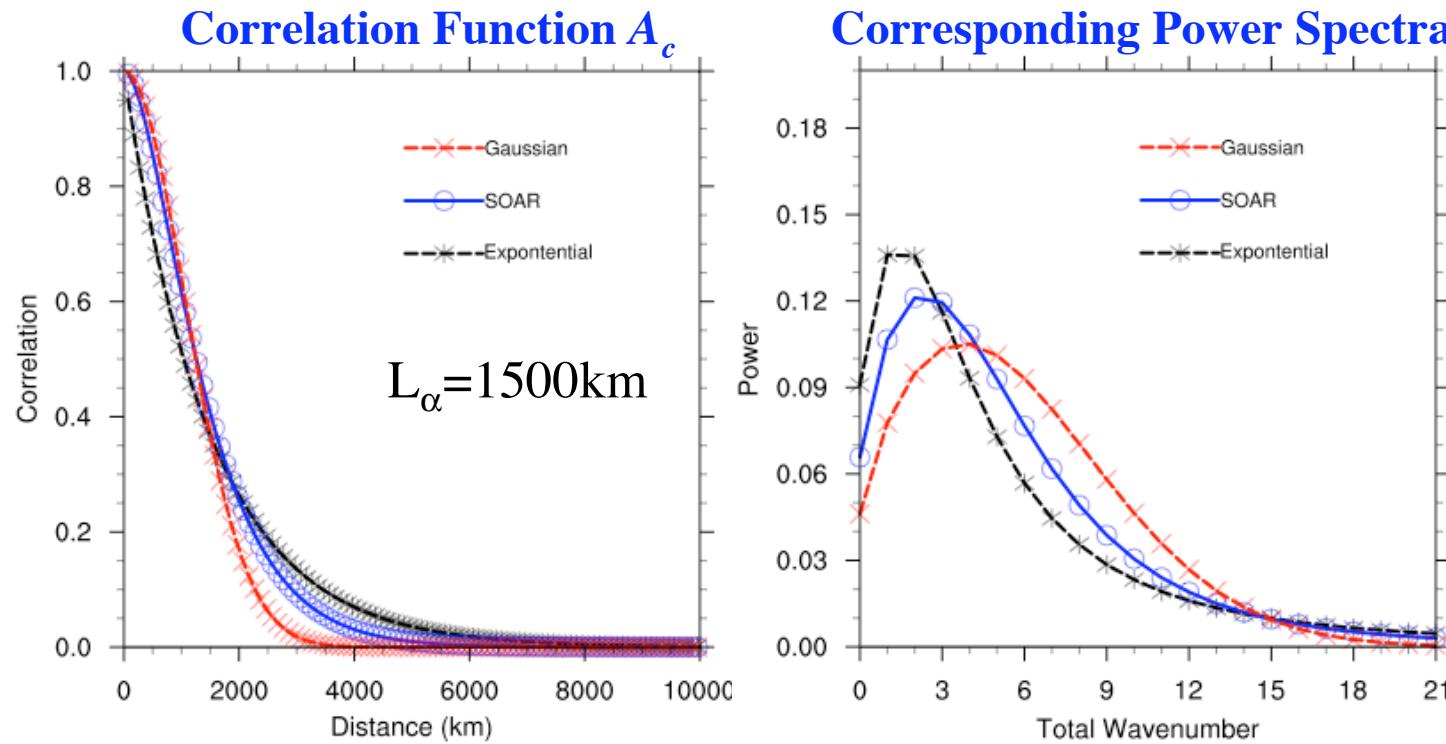
- W_b and W_α are weights defined to conserve forecast error.

- Lorenc (2003)-type hybrid conservation:
$$\frac{1}{\sqrt{W_b}} + \frac{1}{\sqrt{(W_\alpha / \sigma_\alpha^2)}} = 1$$

- Forecast error variance conservation:
$$\frac{1}{W_b} + \frac{1}{W_\alpha / \sigma_\alpha^2} = 1$$

Example Application of ACV in Global WRF-Var

- Alpha correlation A_c is empirical function (e.g. Gaussian) with prescribed scale L_α .
- Lorenc (2003) suggest equivalence between A_c and covariance localization.
- Test ACV approach in global WRF-Var (horiz. correlations in spectral space)

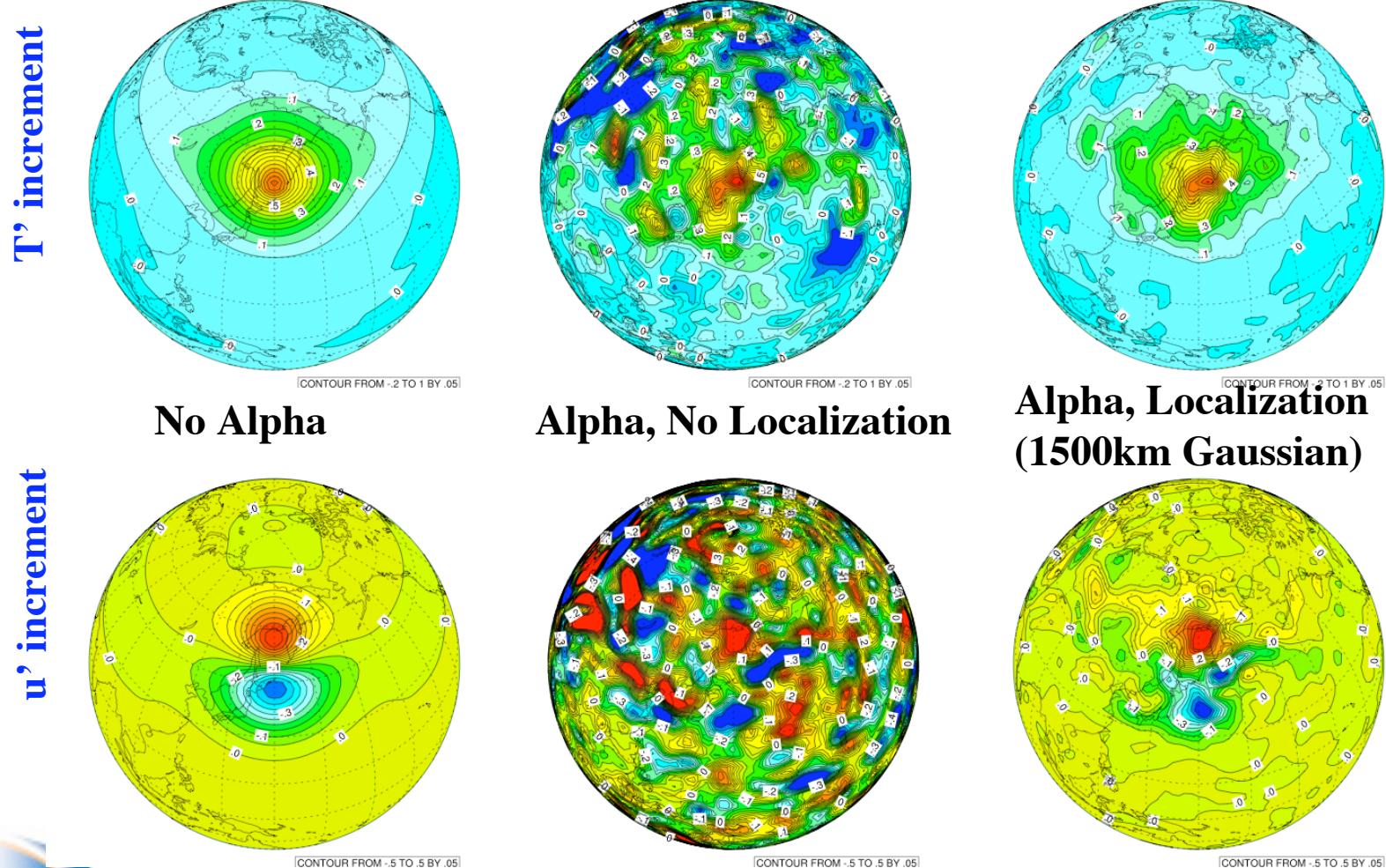


NCAR

Mesoscale & Microscale Meteorological Division

Single Observation Test - Alpha CV

- Specify single T observation (O-B, $\sigma_o=1K$) at 50N, 150E, 500hPa.
- Example: Flow-Dependence given by 1 member of KMA's EPS.

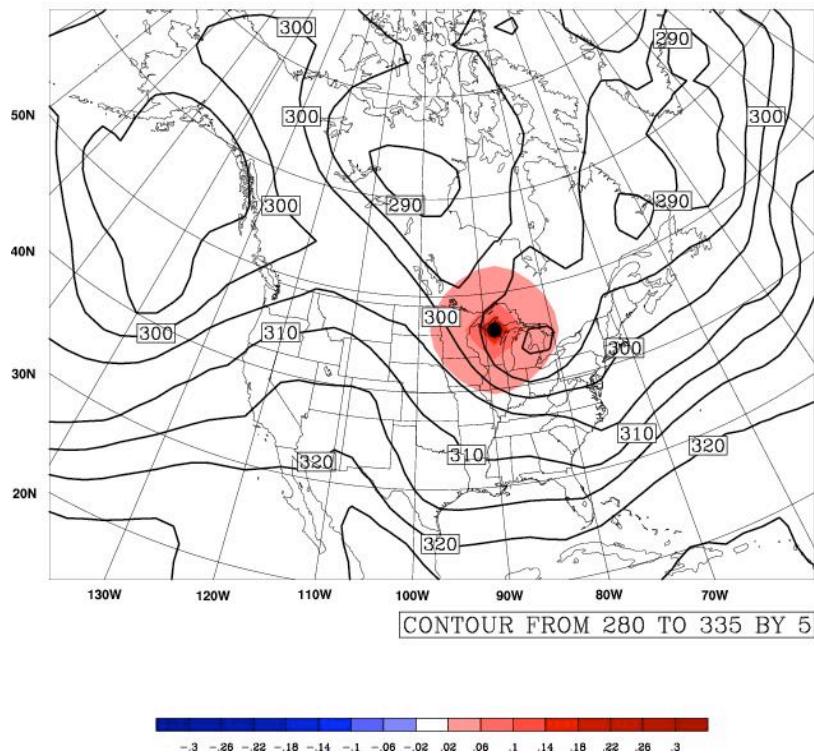


WRF Test with single observation (X. Wang)

Analysis increment

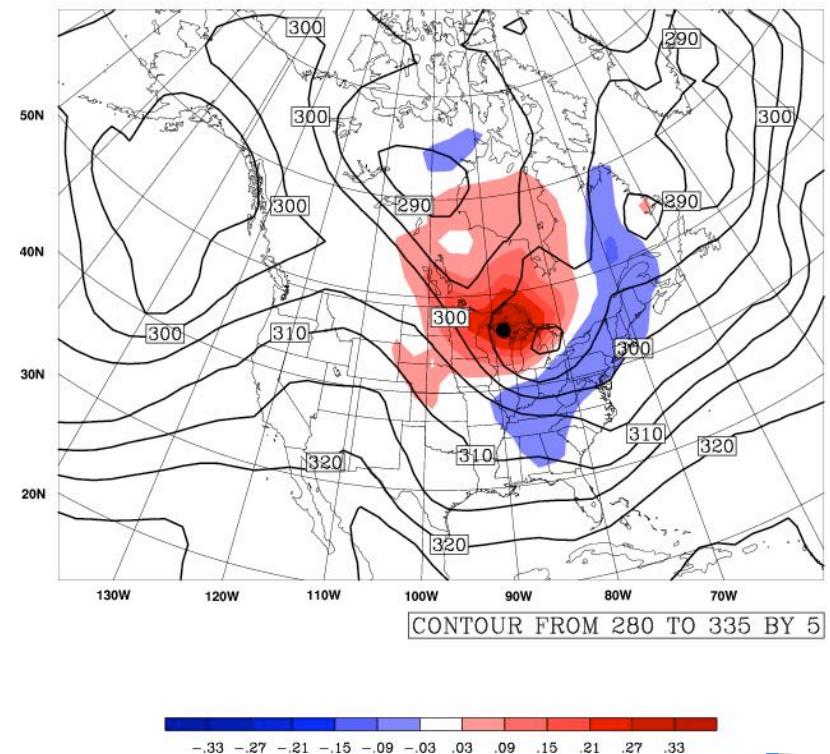
Static covariance

increment of potential T (K) at Level 14 with one T obs at 500mb



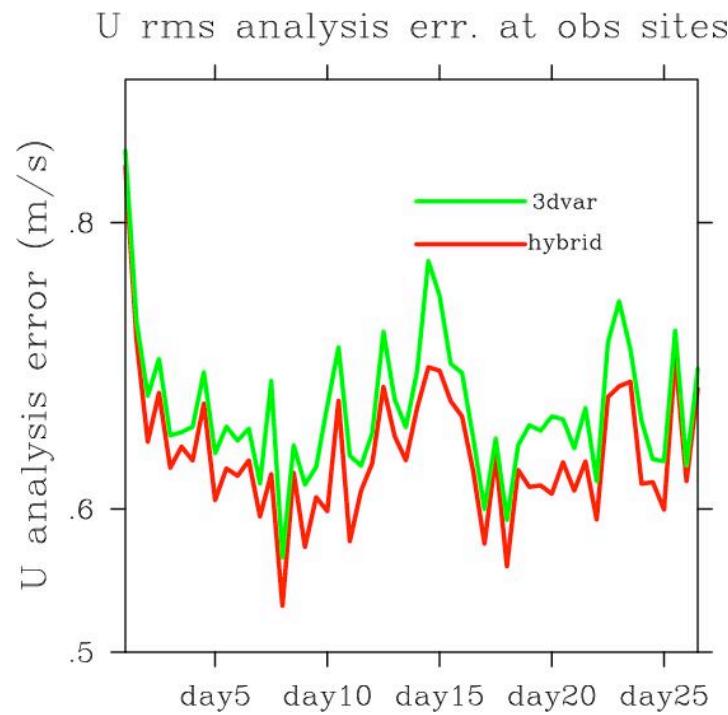
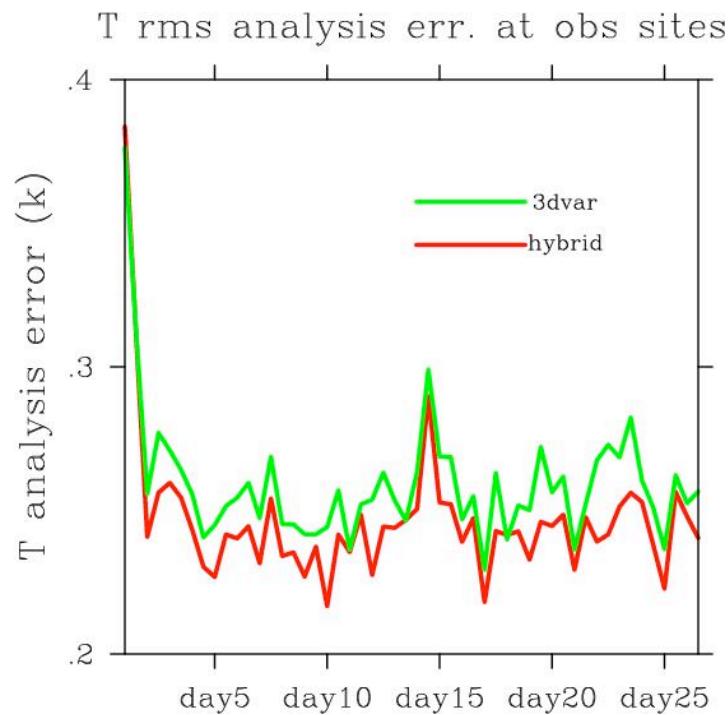
Ensemble covariance with localization

increment of potential T (K) at Level 14 with one T obs at 500mb



Flow-dependent ETKF ensemble covariance is successfully incorporated in WRF-Var

CON200 OSSE Experiment (X. Wang)



- Test hybrid with equal weight (0.5) on static/ETKF error covariances
- Hybrid analyses significantly better than the pure 3DVAR.
- Note yet cycling, nor tuned. Expect further improvements?

Summary

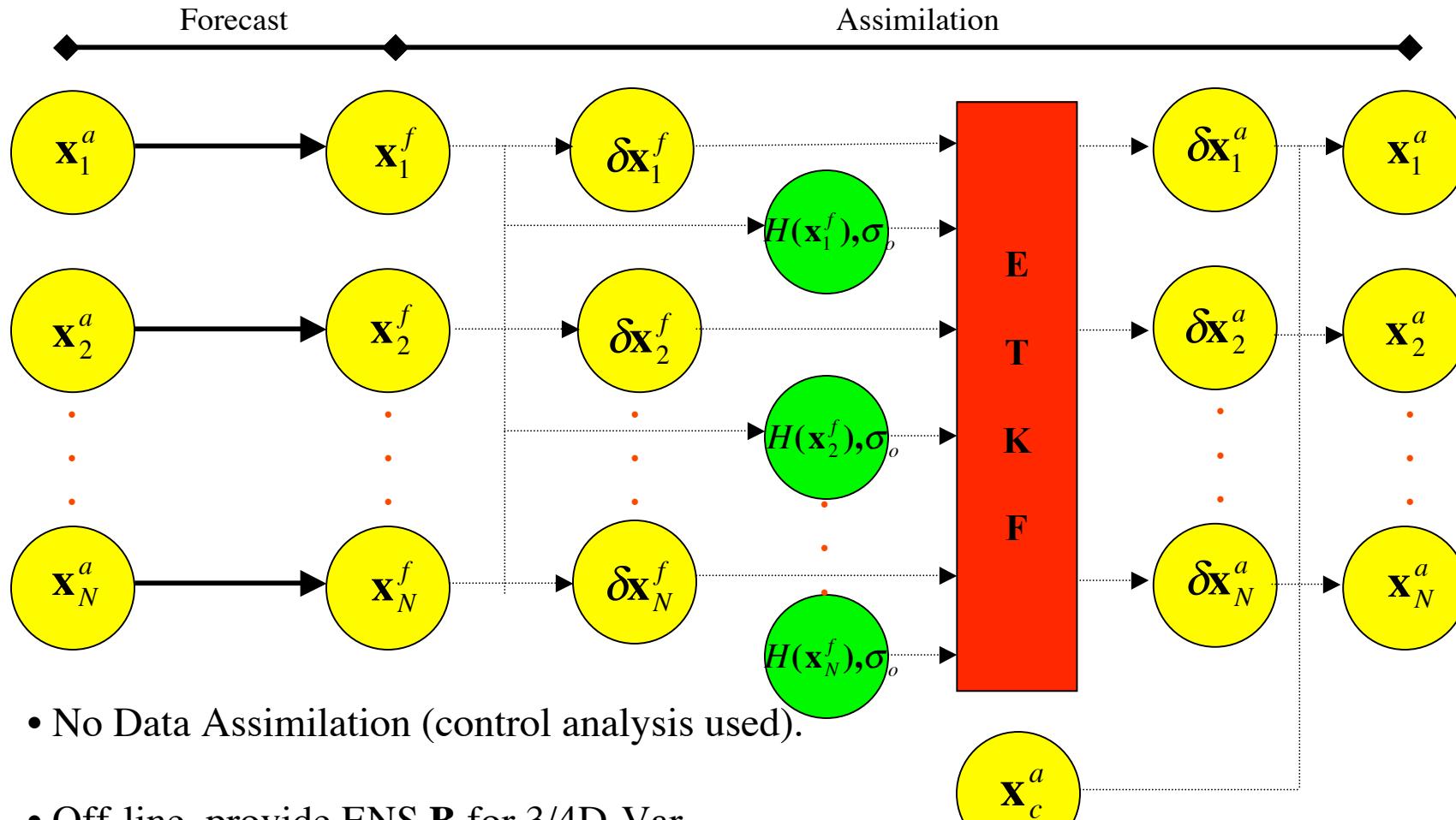
- Accurate estimate of flow-dependent forecast errors likely to be of benefit to well-designed data assimilation system.
- Previous studies indicate hybrid Var/ENS DA superior to pure variational or EnKF for small ensemble sizes. Verified in simple studies here.
- ETKF tested in WRF (including LBC perts.). Adaptive inflation applied.
- Alpha control variable method is an effective way to introduce flow-dependence, even when horizontal correlations in spectral space.
- ACV method is easily coded, has negligible additional cost, is applicable to 4D-Var, and removes spurious sampling error via localization.
- Next stage: Testing and tuning in full NWP system.

Acknowledgements

NCAR's 4D-Var and hybrid DA work is supported by:



Coupled WRF/ETKF EPS



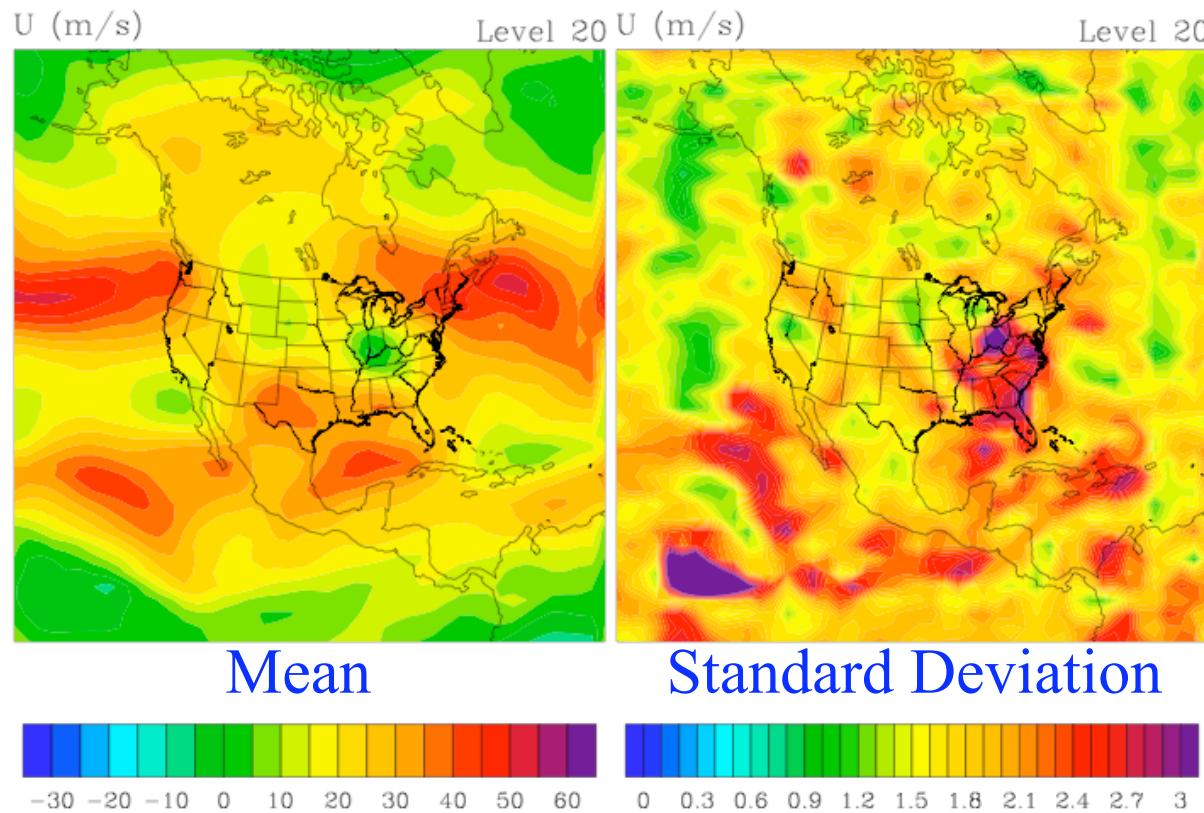
- No Data Assimilation (control analysis used).
- Off-line, provide ENS **B** for 3/4D-Var.
- On-line, provide IC perts for EPS.

$$\mathbf{x}_n^a = \mathbf{x}_c^a + \delta\mathbf{x}_{ETKF_n}^a$$

ETKF Ensemble Mean And Std. Deviation

U Level 20

2003010100 T+12



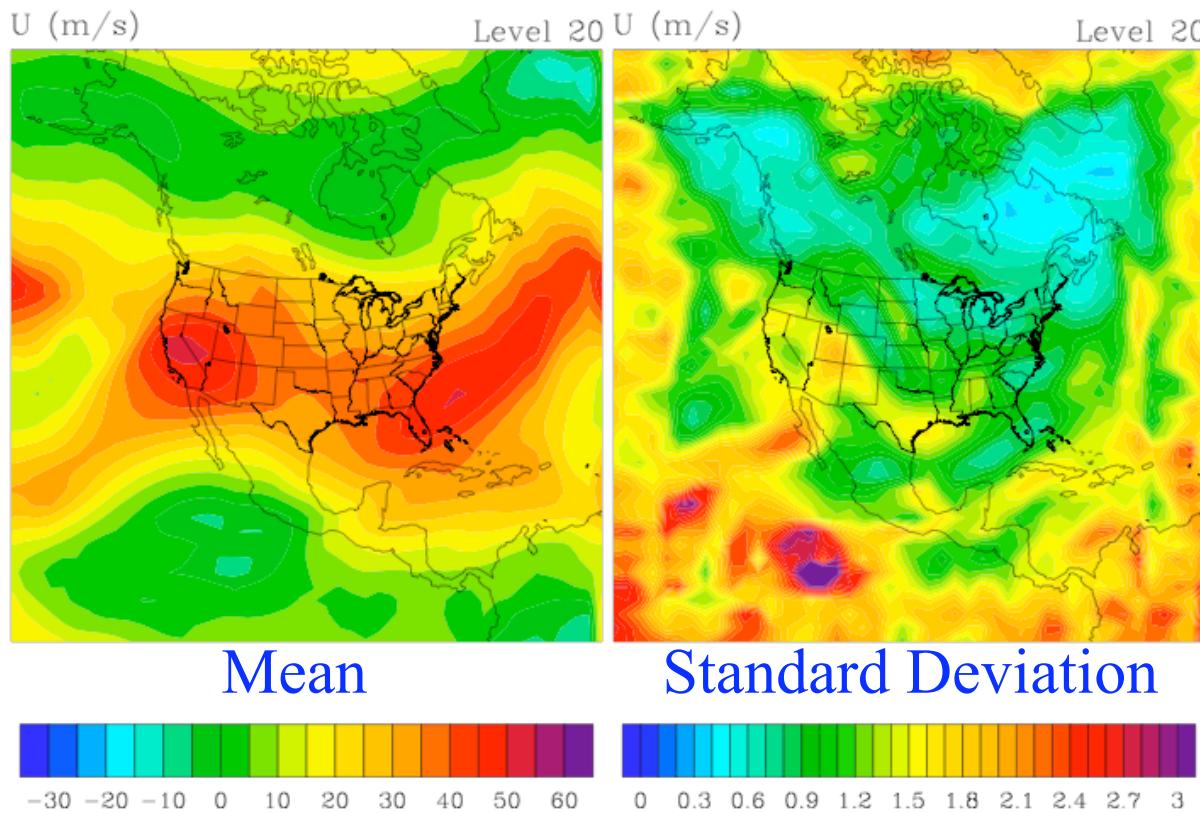
NCAR

Mesoscale & Microscale Meteorological Division

ETKF Ensemble Mean And Std. Deviation

U Level 20

2003011400 T+12

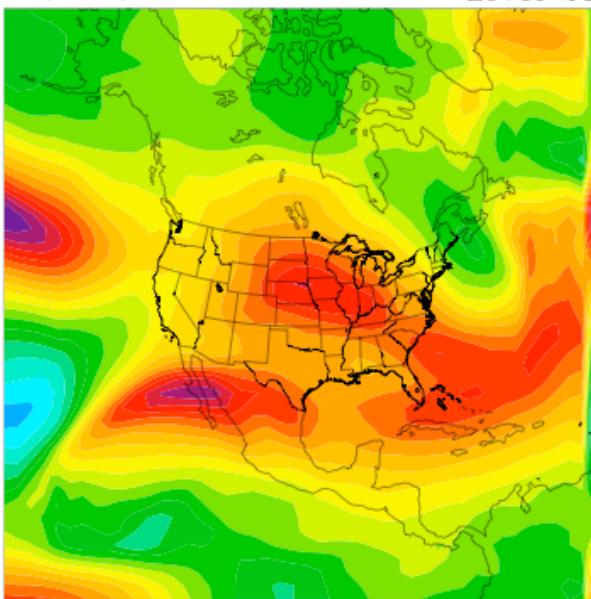


ETKF Ensemble Mean And Std. Deviation

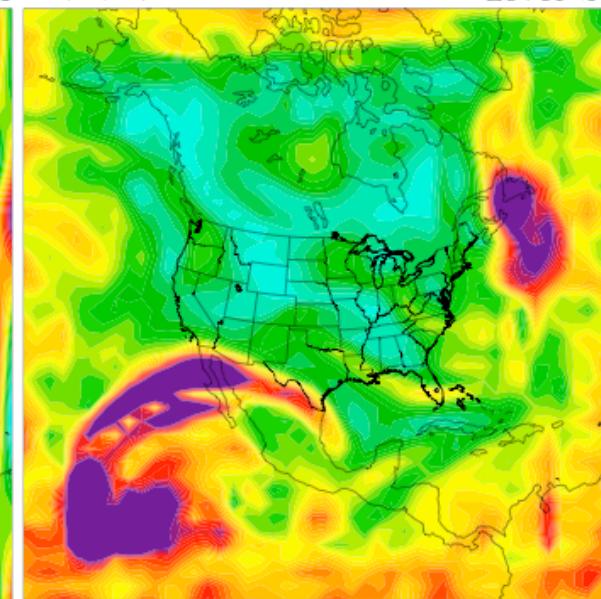
U Level 20

2003012800 T+12

U (m/s)



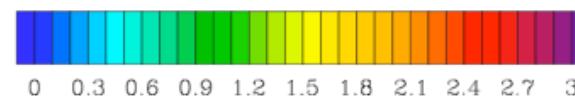
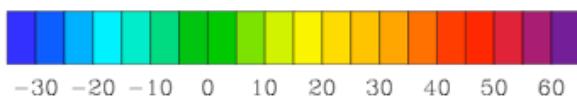
Level 20 U (m/s)



Level 20

Mean

Standard Deviation



NCAR

Mesoscale & Microscale Meteorological Division

Variational Data Assimilation

- The standard components J_b and J_o of the cost function are defined as

$$J_b[\mathbf{x}(t_0)] = \frac{1}{2} [\mathbf{x}(t_0) - \mathbf{x}^b(t_0)]^T \mathbf{B}_o^{-1} [\mathbf{x}(t_0) - \mathbf{x}^b(t_0)]$$

$$J_o[\mathbf{x}(t_0)] = \frac{1}{2} \sum_{i=0}^n [\mathbf{y}_i - \mathbf{y}_i^o]^T \mathbf{R}_i^{-1} [\mathbf{y}_i - \mathbf{y}_i^o]$$

- \mathbf{B}_0 is an *a priori* weight matrix estimating the error covariance of \mathbf{x}^b .
- Direct calculation of J_b and J_o is impossible for NWP problems (\mathbf{B}_0, \mathbf{R} are matrices of dimension $\sim 10^7$). Therefore need to make simplifications.
- Each practical implementation is different!